International Journal of American Linguistics

VOLUME XXVI July 1960 Number 3

ON TYPOLOGICAL INDICES I: RANKING OF LANGUAGES

A. L. KROEBER

UNIVERSITY OF CALIFORNIA

0. Introduction: antecedents
1. Greenberg's basic tabulation
2. Rank orders of values found
3. High rank languages
3.1. Eskimo, Swahili, Yakut, Sanskrit
3.2. Other Indo-European languages
3.3. Annamese
3.4. Language classes provisionally validated and invalidated
3.5. Synthetic trend of Sanskrit
3.6. Check by index summation
4. Intra-Indo-European indices
4.1. Change with period
4.2. Contrary trend
4.3. Isolation and agglutination
5. The value of description by index
6. Possibilities
7. Works referred to

0. In a recent paper in Language concerned primarily with the statistical description and the internal classification of Indo-European, as this subject was recently reopened (also in Language) by Ellegård, I was stimulated by the latter's precedent to add a final section on taxonomy in language study. Not only did I agree with Ellegård that the total field of linguistics needed more systematic classification, but I supported the opinion that we were nearing the point of recognizing that there may have been on the average more external influencing in the history of human languages (diffusion, 'wave theory') than internal genetic development traceable by the standard comparative philological method—much as in the non-linguistic domains of culture. If this is so, no amount of reconstructive phylogeny will give us more than partial and isolated pieces of the total history of human language. With the genetic development of any group of related idioms pursued farther and farther back in time, the evidence becomes more ambiguous and the reconstruction more speculative, until the far end is completely open: the language just has no bottom any longer. So far as we can tell, it is conceivable that all languages now spoken or written on earth go back to a single common origin. But even if so, it is clear that beyond a point which we can already foresee, although we cannot precisely define its time distance, the reconstructed data will be so vague and chaotic that we can never hope to tie them together with assurance. The reconstructed genetic groups will have reduced to narrower and narrower cones and finally mere lines of descent, which, even if they head toward a seeming point of meeting, occupy so little of the total area of speech then existing, that probability will be only infinitesimally small that the vanishing point or points will be near any historic actuality.

In short, the very virtue of the comparative genetic method, the fact that it deals with the strict evidences of point-for-point homologies, also sets a necessary limitation to its scope. What falls outside the cone of its homologies, even though historically it lies within the genetic bundle of idioms investigated, is not reached by the method; while everything else beyond the bundle is of course precluded by definition.

It is thus evident that the main area of linguistics, especially so far as this is concerned with the total multiplicity of languages, will have to be tilled by non-genetic methods. That means non-genetic comparison. At any rate, such comparison, on a

systematic basis, is a first step and a necessary one. It will yield indications of similarity and relationship, whether these be due to internal development or to influence from without; and these indications may suggest what may ultimately be proven to be genetic and what not.

When we reflect on the fact that hundreds of different languages now possess modern, competent descriptive grammars, it is astonishing how little comparative use has been made of them. Either we lack interest in comparing non-genetic groups, or we lack knowledge of how to compare them fruitfully.

Of course comparison is not the only problem in non-genetic linguistics. The purpose of comparison is to establish first the range of variations and similarities, and then to give historic meaning to the synchronic variations and similarities. There is also a synchronic approach which is without ultimate diachronic motivation, but looks for constant principles. This approach has given us, for instance, the concepts of the phoneme and the morpheme. It goes without saying that both interests are legitimate, but they must follow different methods. In the far run, they tend to interact helpfully. For instance, comparisons today can have more meaning than they would have had before the phoneme and morpheme concepts were formulated.

But cross-genetic comparison is out of fashion. Perhaps the presence of the established method of dealing with genetic homologies has helped block its development. I believe that the recent growth of linguistics has been great enough that we are losing opportunities by hanging back from undertaking serious cross-genetic comparison, especially now that the limits of productivity of the genetic method are becoming more evident. It is obvious that we shall never attain to a generic or total taxonomy of language without cross-genetic comparison.

In my paper in Language I suggested that Greenberg's return to a typological approach (in the Wallis Festschrift, see Spencer in the Bibliography) was a promising start, though I also deplored Greenberg's seeming lack of confidence or interest in following up his tentative endeavor. The old deductive and ethnocentric typology of languages that began with the Schlegels and W. (not A.) von Humboldt had come to a dead end with Schleicher. Fifty years later Sapir made a gallant attempt to evolve a sounder and richer typology, and certainly succeeded. Yet in spite of occasional interest by anthropologists and others, the linguistic profession has essentially ignored Sapir's classification. Perhaps this was why he himself never returned seriously to it.

Now, thirty years after Sapir, Greenberg has picked up where Sapir left off, using Sapir's ideas—with one important difference. Sapir still classified languages as unit wholes, according to the degree in which certain structural categories were manifest in them. Greenberg's originality lies in comparing the features or categories instead of the whole languages. This has enabled him to develop a scale for measuring or expressing the degree of development of each feature by a numerical index. The composite typological index is thus an assemblage of indices for ten features. It makes possible the expression of more numerous combinations of features, and represents a definite gain in analytic penetration and flexible description, over what Sapir could attain with his typing of language wholes. Far improved over his predecessors' classes as Sapir's classes of languages were, his synthetic intuitive results were sometimes awkward compromises. For instance, in his Fundamental Type B, "Complex Pure Relational," particular languages are cited as being Agglutinative, Fusional, Symbolic, Agglutinative-isolating, Fusional-isolating, Fusional-agglutinative in 'technique,' each with or without symbolic tinge; and further, the same languages are cross-classified as being also either analytic,

mildly synthetic, synthetic, mildly polysynthetic, polysynthetic.

Greenberg's typological indexing contains, in dealing with features on a quantitative scale for each, a novel point of view and a procedure of broad import. It can probably be extended to include additional features or categories which Sapir did not utilize in his 1921 classification and which therefore Greenberg omitted from his total index. I hope to follow with a paper making some suggestions as to such broadening of scope. Meanwhile, however, it seems worth while to demonstrate that Greenberg's illustration of his index, by application of it to eight sample languages, reveals even more results than he claims for it. Possibly with his quick wits and nimble faculty of generalization, Greenberg has sensed the significances which I am about to make explicit, assuming an audience which would automatically recognize them. Yet the lack of general positive recognition accorded his paper in five years suggests that linguists have construed his casual manner, brevity of substantial evidence, and lack of follow-up, as evidence of a degree of tentative half-heartedness on his own part. I do not share this feeling: the paragraphs that follow aim to show that Greenberg has started something of great possibilities.

1. I reproduce first, as Table 1, Greenberg's tabulation of ten index ratings of eight languages, so that my discussion of them can be more readily followed. It may also be of general convenience to have Greenberg's basic tabulation available in a linguistic journal. His paper is a lone linguistic one in a Festschrift that appears to have had no wide circulation. I use his names for the 10 indices but have renumbered them 1, 2, 3a, 3b, 3c, 4a, 4b, 5a, 5b, 5c instead of 1–10 to indicate their grouping. I have also added the formulaic ratio for computing each index, plus explanation of the abbreviations used in the formulas.

2. Rank orders within each index are substituted in Table 2 for Greenberg's Table 1 values. For instance, under the first index, Synthesis, the Eskimo value of 3.72 is highest, and therefore its entry is 1; Sanskrit is second with 2.59, Swahili a close third with 2.55, and thus down to Annamese with 1.06, the last, or rank number 8. And so on for all ten indices. For convenience, all firsts are blackface, and

TABLE 1

Greenberg's Typological Index Values

		Sk	AS	Per	Eng	Yak	Swa	Ann	Esk
1. Synthesis	M/W	2.59	2.12	1.52	1.68	2.17	2.55	1.06	3.72
2. Agglutination	A/J	.09	.11	.34	.30	.51	.67	..	.03
3a. Composition	R/W	1.13	1.—	1.03	1.—	1.02	1.—	1.07	1.—
3b. Derivation	D/W	.62	.20	.10	.15	.35	.07	—	1.25
3c. Gross Inflection	I/W	.84	.90	.39	.53	.82	.80	—	1.75
4a. Prefixation	P/W	.16	.06	.01	.04	—	1.16	—	—
4b. Suffixation	S/W	1.18	1.03	.49	.64	1.15	.41	—	2.72
5a. Isolation	O/N	.16	.15	.52	.75	.29	.40	1.00	.02
5b. Pure Inflection	Pi/N	.46	.47	.29	.14	.59	.19	—	.46
5c. Concord	C/N	.38	.38	.19	.11	.12	.41	—	.38

M Morpheme; W Word; A Agglutination, lack of morphophonemic change at morpheme juncture; J morpheme juncture; R Root morpheme; D Derivational morpheme; I Inflectional morpheme; P Prefixation; S Suffixation; O word Order; N Nexus, expression of word interrelations; Pi Pure inflection; C Concord.

Based on Greenberg, p. 218, in R. F. Spencer, ed., 1954.

TABLE 2
Rank Orders of Languages Substituted for Index Values

		Sk	AS	Per	Eng	Yak	Swa	Ann	Esk
1. Synthesis	M/W	2	5	7	6	4	3	*8*	1
2. Agglutination	A/J	6	5	3	4	2	1	..	7
3a. Composition	R/W	1	5–8	3	5–8	4	5–8	2	5–8
3b. Derivation	D/W	2	4	6	5	3	7	*8*	1
3c. Gross Inflection	I/W	3	2	7	6	4	5	*8*	1
4a. Prefixation	P/W	2	3	5	4	6–8	1	6–8	6–8
4b. Suffixation	S/W	2	4	6	5	3	7	*8*	1
5a. Isolation	O/N	6	7	3	2	5	4	1	*8*
5b. Pure Inflection	Pi/N	3–4	2	5	7	1	6	*8*	3–4
5c. Concord	C/N	2–4	2–4	5	7	6	1	*8*	2–4

TABLE 3
Frequencies of Certain Rank Orders per Language

	Highest Index	Second Highest	Lowest Index	
			Absol.	Tied
Eskimo	4	—	1	2
Swahili Bantu	3	—	—	1
Sanskrit	1	4	—	—
Yakut Turkish	1	1	—	1
Anglo-Saxon	—	2	—	1
English	—	1	—	1
Mod. Persian	—	—	—	—
Annamese*	1	1	6	1

* 9 indices only.

all eighths italicized; 3–4, 5–8, mean there is a double or quadruple tie for those rank positions.

3. A frequency selection of the probably most significant ranks, the highest and lowest, from Table 2 is given in Table 3 in a rearranged order. It is at once evident that marked differences characterize the languages.

3.1. Eskimo has highest rank 4 times, is lowest or tied for lowest 3 times in the ten indices. It is a language of extremes, evidently, from a generic point of view.

Next comes Bantu Swahili, with 3 firsts, but only 1 tie for last.

Still giving prime consideration to firsts, we find Yakut and Sanskrit with 1 first each. Yakut is tied once for last, Sanskrit's lowest rank is 6.

3.2. The three Indo-European languages other than Sanskrit do not once place first, nor once definitely last, though two of them are tied once for last and each sinks once or twice to next to lowest place. Indo-European is evidently a group not given to extremes. This applies even to Sanskrit, much the oldest of the four Indo-European languages dealt with, for apart from its single instance of rank 1 (due to its fondness for compound nouns), Sanskrit ranges wholly between ranks 2 and 6. In spite of the sample being small, it suggests that Indo-European may prove to be an essentially middle-of-the-road or average type of human language, instead of being the most advanced or perfect as used to be thought.

3.3. Annamese, the only representative in the sample of the long recognized 'isolating' class, comes out, expectably, as being extreme at the opposite end from Eskimo. It achieves 1 first (in 'Isolation'), 1 second (in 'Composition'), and is last or tied for last in its remaining 7 features.

3.4. The upshot is that the Greenberg index-scoring method provisionally validates two of the old classes of languages, but fails to support two others. Isolating and Polysynthetic remain as actual classes; Agglutinating and Inflectional do not. It was in regard to these two latter that ethno-

centric rating was most successfully active in nineteenth-century typing. Indo-European languages were rather remote from obtrusively manifesting either isolating or polysynthetic qualities, so there was less judgment-warping in regard to these two features. To be sure, stem-compounding is undoubtedly a form of synthesis, but the Sanskrit and classic Greek synthetic propensities did not happen to include verb-verb or verb-object compounds nor certain other forms of synthesis, so that it was easy to keep Indo-European out of the polysynthetic 'contamination.'

3.5. It is however clear that Sanskrit, the oldest of the four Indo-European languages subjected to test by Greenberg, has a high overall synthetic propensity, aside from the binary compound nouns in which it ranks number 1. It stands second in rank in no less than four other traits that contribute toward generic synthetic quality, namely, Derivation, Prefixation, Suffixation, and general Synthesis (ratio of morphemes to words).

3.6. This conclusion is confirmed if we summate the values of Table 1. We find the following seriated results: Eskimo, 11.33, Sanskrit 7.51, Swahili 7.06, Yakut 7.02, Anglo-Saxon 6.42, English 5.34, Modern Persian 4.88, Annamese 3.13. Here Sanskrit comes between Eskimo on one side and Swahili and Yakut on the other, and is well separated from the other three Indo-European languages. To be sure, summation is only a rough method of seriating classification, especially when the values of the several indices vary greatly in absolute size—on which point I hope to say more at some later time; nevertheless it probably has some indicative significance.

4. As to the change of structural type that has occurred within Indo-European, the results contained in Table 1 come out more clearly if we isolate the Indo-European index values and then arrange them in some rank order, preferably perhaps the order of strength which the ten traits show in Sanskrit, the earliest. Table 4 gives the results. While Anglo-Saxon is of course ancestral to English, it is so by somewhat less than 1500 years, whereas its time falls about equally long after Sanskrit. It will be seen that in conformity with its midway temporal position, Anglo-Saxon averages about equally intermediate between Sanskrit and the two modern Indo-European languages in its features.

TABLE 4

Indo-European Rank Orders in the Sequence of High Ranks in Sanskrit

Index	Sk	A-S	Pers	Engl
3a. Composition	1	5–8	3	5–8
4a. Prefixation	2	3	5	4
4b. Suffixation	2	4	6	5
3b. Derivation	2	4	6	5
1. Synthesis	2	5	7	6
3c. Gross Inflection	3	2	7	6
5c. Concord	2–4	2–4	5	7
5b. Pure Inflection	3–4	2	5	7
2. * Agglutination	6	5 <	3	4
5a. * Isolation (Word Order)	6	7 <	3	2

4.1. The change in structure from early to late Indo-European is not only consistent but rather drastic. In eight out of ten features, the rank of Sanskrit is in the first half (ranks 1–4) of all the languages considered: these are the first eight indices in Table 4. In six of these eight, Anglo-Saxon also ranks in the upper half, though usually lower (five times lower, once the same, twice higher) than Sanskrit. But as regards these same eight features, both Persian and English seven times out of eight fall in the lower half (ranks 5 to 8).

4.2. In the two remaining features, the last in the table and marked by asterisks, the ranking of Sanskrit is 6, of Anglo-Saxon 5 and 7, but Persian and English are in the UPPER half, Persian in rank 3 and 3, English in 4 and 2. This exceptional direction of change is indicated by the musical sign for *crescendo* (higher index, indicated by rank

order nearer 1) placed between the columns for ancient and for modern Indo-European.

Similarly, the eight majority traits are divisible into two subgroups, the first five and the next three in Table 4. The first five are those most immediately concerned with complexly synthetic words, and in all of these Sanskrit ranks higher than Anglo-Saxon as well as the two contemporary languages. In the next three traits, however, (3c, 5c, 5b), Anglo-Saxon ranks a bit higher than Sanskrit, or is tied with it. These three traits are Gross Inflection, Pure Inflection, and Concord. These have associations with word synthesis but are not directly expressive of it. This statement might be logically challenged for Gross Inflection (I/W); but it holds true at least quantitatively in that inflectional morphemes on the whole constitute lighter or smaller increments (containing fewer phonemes) than derivational or root morphemes or prefixes and suffixes (4a, 4b) as positional overall classes of morphemes.

4.3. On the contrary, the two traits which have been accentuated in Indo-European with the passage of time, as indicated by higher rank order, are Isolation and Agglutination. Isolation refers to frequency of syntactical nexus being achieved through the ORDER of isolated root words. Agglutination is of course not a form of isolation because it refers to a particular kind of combination of morphemes within words. But it indicates combination of the loosest kind, without (or with minimal) morphophonemic change, so that the elements are merely glued together in juxtaposition instead of being welded by morphophonemic process, and therefore are readier to be shaken apart into isolated units. It is thus clear that Table 4 adequately expresses the tendency, long intuitively recognized in Indo-European, and sometimes expressed, for the later phases of these languages to approach the isolating, analytical character of the East Asian languages, and of Polynesian within Malayo-Polynesian.

5. This previous recognition raises the question of why one should bother at all with numerical expression if the phenomena can be recognized by intuitional thought-processes, that is, verbal and non-numerical thinking.

The answer is that when broad problems of comparison of multiple phenomena are undertaken, answers that are both broad and specific can ordinarily be given only as based on some sort of quantitative evaluation. The unaided intuitive mind can carry and balance simultaneously only so and so many variable phenomena, as is necessary in order to ascertain their even more numerous interrelations which underlie the organized classification and then interpretation of the phenomena. The intuitive mind of a conscientious scholar realizes its weakness in such larger situations and refrains from all but the more overwhelmingly patent inductions—from fear of making mistaken judgments where the skein becomes more tangled. The less conscientious investigator falls back on guess work or personal bias or ethnocentrism, and gives answers that contain errors which obscure the situation for later workers. It is the really devoted scholar, unwilling to err and best able to judge, who is most likely to hang back and refrain from broad judgments, particularly in the face of comparative problems. What a set of indices like Greenberg's does is to enable the conscientious scholar to present, by simple counting, arrangement, and arithmetical division, unbiased answers of some degree of probability as to a larger number of interrelations.

It is not in the least a question of pedantry through wearing the false plumes of physical science. There is no algebraic equation used for these indices, not even of the simplest kind—nothing but elementary arithmetic as learned in grammar school.

The intuitive scholarly mind remains in full control of all critical areas: the setting of the problem, the rigging of means, the interpretation of results.

6. I think the foregoing analysis has shown that definitely meaningful results can be got from numerical values for ten indices as applied to only eight languages. If the eight languages were raised to twenty-eight or thirty-eight, including say Greek, Latin, Russian, Arabic, Hebrew, a Hamitic, Dravidian, Indonesian, and Ugric language or two, Japanese, Polynesian, a Melanesian dialect; from native America say a Uto-Aztecan, Mayan, Siouan, Athabascan, Algonkin, and three or four South American representatives, and with several from Africa—what illuminating likenesses and unlikeness, within genetic groups and across them, might not be revealed? And if indices could be devised for twenty instead of ten traits, the qualitative depth and saturation of findings might be doubled or even more than doubled.

As said above, I am planning to return in a future paper to the problem of enlarging the range of Greenberg's typological approach. I hope I have shown here its potential value, even as it stands, for contrastive and constructive classification of languages.

7. Ellegård, Alvar. Statistical Measurement of Linguistic Relationship. Language 35.131–156 (1959).

Greenberg, Joseph H. A Quantitative Approach to the Morphological Typology of Language. Pp. 192–220 of (Wilson D. Wallis Festschrift), ed. by R. F. Spencer, 1954, q.v. [Reprinted in this issue of IJAL, with permission of the University of Minnesota Press. Ed.]

Kroeber, A. L. Statistics, Indo-European, and Taxonomy. Language 36.1 ff. (1960).

Sapir, Edward. Language. Harcourt, Brace and Co., New York, 1921. The classification of languages is in Ch. 6, Types of Linguistic Structure, pp. 127–156.

Spencer, Robert F. Method and Perspective in Anthropology: Papers in Honor of Wilson D. Wallis. University of Minnesota press, Minneapolis, 1954 (Greenberg's paper constitutes pp. 192–220).

A QUANTITATIVE APPROACH TO THE MORPHOLOGICAL TYPOLOGY OF LANGUAGE*

JOSEPH H. GREENBERG
COLUMBIA UNIVERSITY

One of the steps which any science must take if it is to realize the potentialities of the scientific method is to advance beyond mere description to comparison and classification of the objects it studies. That linguistics has taken this step is indicated by the very existence of a subject matter called "comparative linguistics," one which is, moreover, of respected standing among the sciences dealing with man. However, the methods of comparative linguistics represent but one of two fundamental methods by which languages may be compared. The second method, which may be called the typological, is the subject of the present paper. It has had a more checkered career than the historic-genetic method which characterizes comparative linguistics. It is important to distinguish as clearly as possible the differences between the two methods. Each is legitimate in its own sphere, but a confusion between the two methods, as when typological criteria are employed to establish genetic relationships, has had harmful effects in the past.

The genetic-historical method classifies languages into "families" which have a common historic origin. A familiar example is the descent of the modern forms of Romance speech, French, Spanish, Portuguese, Italian, Rumanian, etc., from an original unified Latin as the result of changes of the same speech form in distinct areas. When this diversification has not been so remote in time as to remove all evidential traces, comparison will reveal characteristic resemblances among languages which have such a common origin. What is chiefly relevant for such comparisons is the resemblance between individual forms in language, both with respect to sound and meaning. For example, English *nose* and German *Nase* have similar sounds and both have virtually identical meanings, "*nose*"; English *hound* and German *Hund*, "dog," are similar in sound and similar though not identical in meaning. Any language consists of thousands of forms with both sound and meaning, the relation between the two being arbitrary. In principle, any sound whatever can express any meaning whatever. Therefore, if two languages agree in a considerable number of such items, as do German and English, and if—a problem not discussed here—the resemblances cannot be explained by borrowing, we necessarily draw a conclusion of common historic origin. Such genetic classifications are not arbitrary, in the sense that there is no room for the establishment of varied criteria leading to different results. This is because such classifications reflect historic events, which must either have happened or not have happened. Either the speakers of German and English speak a language transmitted from an original unified proto-Germanic speech community or they do not. The analogy here to biological classification is extremely close as, indeed, the use of such terms as *family*, *related*, and *genetic* might suggest. Just as in biology we classify species in the same genus or high unit because the resemblances are such as to suggest a hypothesis of common descent, so with genetic hypotheses in language.

One can, however, compare languages which cannot be shown to be related

* This paper first appeared in the Festschrift for Wilson D. Wallis, published by the University of Minnesota Press, *Method and Perspective in Anthropology: Papers in Honor of Wilson D. Wallis*, edited by Robert F. Spencer (1954).

genetically and this either with respect to some features of sound only or meaning only. The two following examples will serve at once to illustrate this possibility and to show that, in this way, legitimate scientific problems arise. All languages must express comparison—the fact, for example, that one thing is larger than another. If we compare all the languages of the world in this respect, we discover that the number of methods is limited—e.g., a special, inflected form of the adjective (English *greater*), the use of a preposition meaning *from* (Semitic), the use of a verb meaning *surpasses* (as widely used in Africa)—and that of these some are far more frequent than others and have definite geographical distributions which disregard genetic boundaries. All these are surely data of real interest which require an explanation. The phenomenon just discussed is semantic. In the area of sound patterns, the independent appearance of a system of five vowels with two degrees of length—a, a·, e, e·, i, i·, o, o·, u, u·—in classical Latin, Hausa in West Africa, Yokuts, an Indian language of California, and doubtless elsewhere—is surely worthy of attention. Indeed, such phenomena have been studied by Trubetskoy and others in an attempt to discover what types of vowel and consonant systems are possible and their relative frequency and area of appearance.

If genetic comparison establishes classes of languages, i. e., language families in the accepted sense, does not typological classification do this also? The answer, of course, is that it can but that, as opposed to genetic classification, it has no specific historic implications and is arbitrary, i. e., will lead to different results depending on the criterion or combination of criteria selected.

In this respect it is like racial classification based on a number of arbitrarily selected traits. If, for example, we select such a purely phonetic criterion as the presence or absence of lip rounding as a feature distinguishing pairs of vowel phonemes, the languages of the world will fall into two groups, those that avail themselves of this particular principle of contrast and those that do not. English and Italian, which do not, will fall in class A with innumerable other languages; and French and German, along with a smaller number of other languages from different parts of the world, will fall into class B. If some other typological trait is selected, say the position of the dependent genetive relative to the noun, again two classes of languages will emerge, but they will not coincide with those obtained by the criterion of rounded versus unrounded vowel. Taking both factors into consideration, there will be four classes. Certain characteristics will allow of more than two classes. If languages are classified semantically on the basis of their numerical system, we shall obtain a number of classes of languages: binary, quinary, decimal, duodecimal, and no doubt others as well. In short, in contrast to genetic classification, the number of language groups and their membership will differ with the number and particular selection of linguistic traits utilized for the comparison. At one extreme, if some such feature as the presence of a vowel system is employed as the sole criterion, the languages of the world will fall into two groups, those with (and this group will include all the languages of the world) and those without. This latter class will, of course, have no member. At the other extreme, one can specify so many traits that each language becomes the only member of a specific type.

Many such classifications, as should be evident from the examples just cited, are not very useful. We seek to establish typologies which involve characteristics of fundamental importance in language and which are useful for a variety of reasons. Such a classification does exist, the nineteenth-century division of languages which, in its classical version, was threefold: isolating, agglutinative, and inflective types.

This *trop fameuse classification*, to use Meillet's trenchant phrase, betrays fundamental weaknesses which have inevitably led to its present state of disrepute. Yet the problem itself *is* of sufficient importance for Sapir, in the only major contribution to the subject since the nineteenth century, to have made it the central topic of his book *Language*. However inadequate the nineteenth-century discussion of this topic now appears, the major virtue of the schemes advanced remains. Something of fundamental importance to the over-all characterization of a language, the morphological structure of the word, was instinctively seized upon as the basis of classification, and Sapir's contribution simply continued this essential feature of the earlier approach, in revised form. Other typologies, particularly phonological, are possible, and these latter have been the chief center of interest in typological discussions in recent years. However, the problem of a morphological typology remains, as evidenced by Rulon Wells's recent statement, "It is shocking that there is no established taxonomy of the languages of the world" (1950, p. 31). The revived interest in this topic is further evidenced by the virtually simultaneous and independent development by Charles Hockett of Cornell and myself of the ideas presented in this paper.[1] In short, the time seems propitious to re-examine the nineteenth-century approach to the problem, discarding what the intervening period of linguistic criticism has demonstrated as invalid and incorporating recent methodological advances in order to reformulate the hypotheses along more rigorous lines. A brief critical review of earlier attempts may serve to place the present treatment in its proper perspective.

[1] The same basic approach was presented by Charles Hockett in an unpublished paper read at the annual meeting of the Linguistic Society of America, Philadelphia, 1949, and in a talk I gave before the Linguistic Circle of New York at Columbia University, January, 1950.

The germ of all later classifications is found in the distinction, first set forth by Friedrich von Schlegel in his essay "Ueber die Sprache und Weisheit der Inder" (1808), between languages with affixes and languages with inflections. The valuational attitudes so prominent throughout the later history of the theory are already present in this earliest formulation. The affix-languages express relations through a merely mechanical process. In a striking figure they are likened to a "heap of atoms which every wind of chance scatters or sweeps together" (p. 51). Only Indo-European languages are inflectional, but Semitic is tending in that direction.

This twofold classification was elaborated into a tripartite one by Friedrich's brother, August von Schlegel. In an essay "Sur la littérature provençale" (1818), Schlegel describes three classes of languages: "languages without grammatical structure, affixing and inflectional languages" (p. 559). Of his first class, called by later writers isolating or root languages, he says, "One might say that all their words are roots, but sterile roots which produce neither plants nor trees" (p. 159). To create a literature or science in such a language is a *tour de force*. The affixing languages use added elements ("affixes") to express the relations and nuances of the root ideas, but these affixes still have an independent meaning. Regarding inflectional languages in which such affixes are meaningless (i.e., devoid of concrete meaning), we learn that to them can be attributed a kind of organic life ("organisme") in that "they contain a vital principle of development and growth" (p. 159). As with later writers, Chinese is his model of a root-language; only the Semitic and Indo-European languages belong to the inflectional group; all the others belong to the vast and heterogeneous intermediate or agglutinative class. The inconvenient fact that the Indo-European languages have tended to lose their inflection bothers him as it did other later writers. He

therefore introduces a further subdivision of his inflectional languages, the earlier synthetic and later analytic. All analytic languages we know result from the decomposition of synthetic languages. Passing over other writers who discussed the topic in much the same terms as Schlegel, we come to Wilhelm von Humboldt who, in his essay "Ueber die Verschiedenheit der Menschlichen Sprachen" (1836), placed this type of analysis at the very heart of his approach to language. Von Humboldt viewed each language as a distinct self-revelation of the spirit (*Geist*). Such self-revelations, while each a valid expression in its own right, exhibit lesser or greater degrees of perfection. There are four classes of languages in von Humboldt's scheme. He adds a fourth, incorporating type to the by now traditional threefold classification in order to accommodate certain American Indian languages whose very complex word-patterns include instances in which the object of a verb is incorporated in the same word as the verb root. Von Humboldt is explicit in rejecting any historical evolutionary interpretation in which higher types evolve out of lower types. These are ideal types involving different degrees of the unfolding of form. The isolating languages are "formless," the incorporating languages, through their over-elaboration, betray no true sense of form. As might be expected, only the inflected languages, by their harmonious fusion of root and affix in a true unity, are credited with a true sense of form.

The definitive exposition of the theory is found in the writings of A. Schleicher, who stood under the twin influences of Darwin and Hegel. The classes of languages are now interpreted as so many historical evolutionary phases of the development of languages. The types are limited to three, which are equated to the stages of the Hegelian dialectic. The inflectional class of languages becomes a higher synthesis arising out of a previous opposition. The decay of inflection in historic times introduces a new phase of the *Geist* in which the material side of language no longer counts. In accord with the intellectual currents of the age, and equipped with an impressive set of quasi-algebraic formulas to indicate the various relations of the root to subordinate elements, Schleicher's version of the theory won wide acceptance and provided the basis for the exposition by those two great popularizers of linguistic science, Max Mueller in Europe and Dwight Whitney in the United States. Later versions such as Steinthal-Misteli, which added to the complications of the scheme without any compensating advantages, never became as popular as Schleicher's, which thus established itself as the basic form of the theory.

During this whole period criticism of the ethnocentrism and vagueness of these typologies was by no means lacking. To cite but one instance, Whitney, who was by no means as enthusiastic about the Schleicherian typology as his European contemporary, Max Mueller, declares "*loved* from *love* is as good a preterite as *led* from *lead* or *sang* from *sing*" (1876, p. 362). *Loved* is, of course, an example of agglutinative technique, while *led* and *sing* are inflective. Still, in regard to isolating languages he talks of "this lack of resources possessed by more happily developed languages; . . . thought is but brokenly represented and feebly aided by its instrument." Others, particularly at a later date, were highly critical or even contemptuous as, for example, Mauthner, according to whom " . . . the valuation [of languages] according to whether their inflections are more or less transparent is as foolish as if one judged the merit of European armies according to the greater or lesser visibility of their trouser seams" (1923, p. 309).

Sapir's treatment of the topic in his book Language (1921) marks an epoch. He firmly rejects both the valuative and evolutionary aspects of the theory. There

is no real reason to suppose that Chinese or Hungarian is not as effective an instrument of thought as Latin or English. "When it comes to linguistic form, Plato walks with the Macedonian swineherd, Confucius with the head-hunting savage of Assam" (p. 234). Man must have had language for at least 500,000 years; hence, if there is indeed a line of development—isolating, agglutinative, inflective—present-day isolating languages cannot possibly represent the primitive stage. Indeed, the evidence of human paleontology and glacial geology had made this assumption long since untenable. Chinese, the classic instance of an isolating language, was known by Sapir's time to have formerly possessed a more complex morphological system, both from earlier records of the language and comparison with Tibetan and other related languages.

Perhaps more significant than these no longer tenable assumptions were other logical defects which had been pointed out from time to time. The distinctive criteria of the various types had never been either defined with clarity or applied with objectivity. In reading Steinthal-Misteli, one has the feeling that the writer is playing with loaded dice. Whenever, on the basis of his own showing, a non-Aryan, non-Semitic language appears to have a praiseworthy feature, a tacit shift in definition shows that "true form" or "true inflection" is not involved. The definitions were not only vague but partly referred to quite different things, so that a language might well belong to several of the supposedly mutually exclusive classes simultaneously. For example, agglutination was usually taken as referring to a technique of mechanical affixation, so that, in Max Mueller's words, "The difference between an Aryan and Turanian is somewhat the same as between good and bad mosaic. The Aryan words seem made of one piece, the Turanian words clearly show the sutures and fissures where the small stones are cemented together" (1890, p. 292). The contrary term should be *inflection*. But the term *inflection* was also used to indicate the presence of affixes without concrete meaning to denote relations among words of the sentence; e.g., by case endings in the noun or person-number terminations in the verb. On this basis Turkish is both agglutinational (on the basis of technique) and inflectional (because of its case and verb conjugational systems).

A further defect, as was pointed out by some critics, was that a language had to be assigned to a single category, although the features employed might be present to a great or less degree. A term like *agglutinative* applies primarily to a single construction. A language may well and indeed usually does contain some agglutinational as well as some nonagglutinational constructions. In other words, it is a matter of over-all tendency rather than absolute presence or absence of the diagnostic traits. Sapir, by distinguishing among the various criteria which were unconsciously being employed in a confused way in the classical theory, constructs a more complex system in which languages are classified by a number of independent criteria and in which the traditional terms are retained, but in well-defined uses which are often on different axes so that they are no longer mutually exclusive.

One such axis distinguished by Sapir may be said to relate to the gross complexity of the word, i.e., the degree of complexity exhibited on the basis of the number of subordinate meaningful elements it contains. The terms employed here by Sapir are *analytic*, *synthetic*, and *polysynthetic*, in ascending order of complexity. The theoretic extreme of analysis is represented by languages in which each word consists of only one meaningful unit and thus has no internal structure. Languages which actually approach this extreme are Chinese, Annamite, and Ewe (in West Africa). These are the languages traditionally called isolating, but, as we shall see, Sapir reserved this term for

another use. Languages like English with words of little complexity were included by Sapir in the analytic group. However, the degree of synthesis is but one criterion and a relatively superficial one, since it does not tell us in what the complexity of the word consists. A second and quite different consideration refers to technique of construction. The contrast is, roughly, between languages in which the subordinate elements are added to the root elements mechanically, i.e., without either of the elements being modified (the most common meaning of agglutination in the classical scheme), and those involving a process of fusion by which the constituent elements become difficult to recognize and separate. To quote Sapir's examples: *good* + *ness* in English is agglutinative, *dep* + *th*, fusional. After treating this topic several times with varying conclusions, Sapir finally sets up a fourfold division: *a*, isolating; *b*, agglutinative; *c*, fusional; *d*, symbolic. Isolation, by which Sapir means significant order of elements, is included here because he thinks of this rubric as concerned with techniques of relating elements. Just as we know that in "John hit Bill" "John" is in construction with "hit" as the subject of the verb by the fact that it precedes the verb, so in *dep-th*, the modification of *deep* to *dep-* indicates that it is in a construction with *-th*. Since the agglutination-fusion contrast refers merely to the mechanics of expression rather than to what is being related, Sapir considers this scale likewise somewhat superficial, though useful as a supplementary criterion.

The division which appears to Sapir the most basic rests on the following considerations. There are two types of concepts which all languages must express, a stock of roots with concrete meanings, e.g., *table*, *eat*, and pure relational ideas which "serve to relate the concrete elements of the proposition to each other," thus giving it definite syntactic form; e.g., Latin *-um* as a mark of the verb object. As constituting the two extremes of concrete and abstract, Sapir puts these two classes at the ends of his scale as I (concrete), IV (pure relational). Between these, he places two groups of concepts which are dispensable, since some languages have them and others do not. Class II consists of derivational concepts which "differ from type I in expressing ideas irrelevant to the proposition as a whole but that give a radical element a particular increment of significance and that thus are related in a specific way to concepts of type I." Sapir's example is the suffix *-er* in *farmer*, which gives an increment of meaning to *farm-* but is not concerned with the structure of the rest of the sentence. A language which does not have concepts of type II would employ a single unanalyzable element to convey the meaning *farmer*. Concepts of type III (concrete relational concepts) lead over to type IV (pure relational), insofar as they help to relate members of a sentence to each other, but differ in that they have an element of the concrete in their meaning. Examples are the elements indicating gender in a language like German. The *-er* of *d-er* in "der Bauer tötet das Entelein" (the farmer kills the duckling) relates *d-* to *Bauer* by agreement in number, gender, and case. It thus indicates that *d-* modified *Bauer* and that *Bauer* is singular and subject of the sentence. It does this, however, with the material alloy of the indication of sex gender, in this case, masculine.

Such concepts are called by Sapir concrete-relational. Since we have four types of concepts, of which I and IV are necessary in all languages, but II and III are dispensable, we set up the following four classes of languages: A consists of those with I and IV only. These Sapir calls Simple Pure-Relational Languages, e.g., Chinese. Group B contains those with II as well as the indispensable I and IV. These are Complex (i.e., Deriving) Pure-Relational Languages. Group C is constituted by languages which have I, III, and IV, but not II. These are

Simple Mixed-Relational Languages. In group D, finally, we have those languages which have all four types of concepts, i.e., Complex Mixed-Relational Languages. Of the two considerations, the presence or absence of II (derivational) or presence and absence of III (concrete-relational) Sapir considers the latter, which aligns types A and B versus C and D, as the more fundamental. In his final table of classifications, Sapir takes a number of languages and, employing the sets of criteria already discussed, first assigns a particular language to one of the "fundamental types" A, B, C, and D just mentioned. He likewise indicates degree of synthesis.

The third factor, technique, is specified separately for each of the groups of concepts II, III (when they are present), and IV, using the above mentioned scale: *a*, isolating; *b*, agglutinative; *c*, fusional; *d*, symbolic. Sapir often mentions two and sometimes three techniques. If weakly developed, he encloses the symbol in parentheses. He then gives an over-all estimate of the dominant technique of the language, often using compound terms such as agglutinative-fusional to indicate roughly equal use of these two techniques. Note that Sapir does not use the term *inflectional* anywhere in his finished classification. He defines *inflection* as the use of fusional techniques in the sphere of inflectional units. He believes that, so defined, it is not of sufficient importance to figure as a fundamental term in his classification. The presence of inflection is, therefore, indicated by the appearance of *c*, indicating fusion, or *c* and *d*, indicating fusion and symbolism, in association with the concepts of group III (mixed-relational). As an example of Sapir's total scheme, let us take his classification of Semitic. These languages are assigned as a group to D, i.e., complex-mixed relational languages with all four types of concepts present. They are synthetic. In the area of derivational concepts (II) the techniques are listed as *d*, *c*, in that order, i.e., symbolic and fusional. In the area of mixed-relational concepts (III) the techniques are given as *c*, *d*, fusional, symbolic. Under IV the technique is (*a*), isolation, i.e., significant word order, the parentheses indicating its weak development. There is finally an over-all estimate of technique as symbolic-fusional.

The present method utilizes the Sapir classification in revised form. The basic criticisms of Sapir, already voiced by Most (1948, pp. 183–190), boil down to two. The first and most significant is that in his classification into four basic types, Sapir seems to be talking about concepts, but in reality his test is a formal, not a semantic one, a fact which leads to some difficulties in exposition. For example, Sapir discusses the concept of the plural, which he considers highly abstract. However, it can be assigned anywhere along the scale I-IV in a particular language, as he points out. His test, then, as to whether the plural is a root concept (I), a derivational one (II), or relational (III or IV) depends upon which formal class a particular language assigns it to. Sapir recognizes this disparity. "It is because our conceptual scheme is a sliding scale rather than a philosophical analysis of experience that we cannot say in advance just where to put a given concept" (1921, p. 117). In the typology presented here, the approach is formal. It is recognized that there is, in fact, a tendency for meanings of root morphemes (Sapir's I) to be more concrete in meaning than derivational (Sapir's II) or inflectional morphemes (III or IV), but this is too vague to constitute a valid procedure. Here, as elsewhere in modern linguistics, we isolate our distinctive units by a formal, not a semantic, test for purely practical reasons. The second criticism relates to Sapir's scale: *a*, isolating; *b*, agglutinative; *c*, fusional; *d*, symbolic. Isolating is a technique of relating, as are the other devices, but applies almost only between words, relative order being rarely significant within the word. It is therefore out of place

here, and this betrays itself in the asymmetry of its occurrence in Sapir's scheme, for it only appears as a technique under IV (pure relational concepts) and does not refer to relations within the word as with the other techniques, but between words.

The method of classification proposed here is fundamentally that of Sapir, with certain modifications in the light of these criticisms. Moreover in place of intuitive estimates based on over-all impressions, an attempt is made to define each feature involved in this classification in terms of a ratio of two units, each defined with sufficient rigor and by the calculation of a numerical index based on the relative frequency of these two units over stretches of text. Five bases of classification are set up in place of Sapir's three, and a set of one or more indices are established to measure the place of any language in regard to each of them. The first of these parameters is the degree of synthesis or gross complexity of the word. Since Sapir's time the minimum meaningful sequence of phonemes in a language has come to be called the morpheme in American linguistics. For example, the English *sing-ing* contains two morphemes but forms one word. The ratio M/W where M equals morpheme and W equals word, is a measure of this synthesis and may be called the synthetic index. Its theoretical lower limit is 1.00, since every word must contain at least one meaningful unit. There is no theoretical upper limit, but in practice values over 3.00 are infrequent. Analytic languages will give low results on this index, synthetic higher, and polysynthetic the highest of all.

The second parameter refers to technique. At one extreme are languages in which the meaningful elements are joined with slight or no modification. This is the classic meaning of *agglutination*. The opposite phenomenon is reciprocal modification or merging of the elements. Actually a number of types of constructions can be distinguished, and thus a more subtle typology might be developed. For purposes of the present paper an alternative is selected which seems to correspond most closely to the intent of Sapir and of the usual nineteenth-century analyses. To use contemporary terminology, what is involved is the degree of morphophonemic alternation. The meaningful stretch actually found in an utterance is called a "morph." A number of related morphs are subsumed under a single fundamental unit, the morpheme. The various morphs are then said to be in alternation. For example, in English we relate the morph *lijf* ("leaf") with the morph *lijv-* which only occurs with the plural morph *-z* to form *lijvz* ("leaves"). *Lijf* and *lijv-* are morphs which alternate within the same morphemic unit. The rules for the statement of this alternation belong to the morphophonemic section of the description of English. Where there are no variations among the constituent morphs of a morpheme or where the variations which exist are all automatic, the morpheme itself is said to be automatic. By an automatic alternation is meant one in which all the alternants can be derived from a base form by a set of rules of combination that holds in all similar instances throughout the language. This matter will be discussed in greater detail below. If both morphs in a construction belong to morphemes which are automatic, the construction is called agglutinative.

The index of agglutination is the ratio of agglutinative constructions to morph juncture. There is necessarily always one less morph juncture in a word than the number of morphs. Thus *leaves* has two morphs but one morph juncture. The index of agglutination is A/J, where A equals the number of agglutinative constructions and J equals the number of morpheme junctures. A language with a high value for this index will be agglutinational, one with a low value fusional. In general, the lower the first, or synthetic, index, the fewer the morph junctures which occur and the less the

importance of this second index in characterizing the language. If a language reaches the theoretic lower limit of the synthetic index 1.00, this second index becomes impossible of calculation because no morpheme junctures occur; that is, the agglutinational index becomes 0/0, which is meaningless. In calculating this index the differences between the degree of agglutination which might be found in constructions involving Sapir's concepts of groups II, III, and IV and which, as we have seen, figures in Sapir's final formulation, have been ignored. Such indices might be calculated on the basis of the distinctions between root, derivational, and inflectional morpheme classes described below, since it is these features which correspond most closely to Sapir's division of concepts. They were not set up, partly in order to avoid too great over-all complications in the typology, partly because of the considerable labor involved in their calculation.

The third parameter corresponds most closely to what for Sapir was the most fundamental basis of classification in languages, the presence or absence of derivational and concrete-relational concepts. Since, as has been seen, an approach in terms of the meaning of concepts is too vague to handle with any degree of rigor, the present treatment is based on the possibility of the exhaustive division of morphemes into three classes, root, derivational, and inflectional. Every word must have at least one root morpheme and many words in many languages have no more. The existence of more than one root morpheme in a word is called compounding. This is a significant feature in which languages differ from each other considerably. There are languages which cannot compound at all, or only very sparingly. Others again compound freely, while most languages are somewhere between. It is remarkable that Sapir nowhere seems to take this into account in his typology. It can easily be measured by a compositional index R/W, where R equals number of root morphemes and W equals number of words. The second class of morphemes is the derivational. Examples of derivational morphemes in English are *re-* in *re-make*, *-ess* in *lion-ess*, *-er* in *lead-er*. The derivational index is D/W, the ratio of derivational morphemes to words. Languages with a high D/W will belong to Sapir's Complex or Deriving subtypes and will thus fit into classes B or D in his scheme. The inflectional morphemes are the third class. Examples in English are *-s* of *eats* and the *-es* of *houses*.

The inflectional index is I/W, i.e., the proportion of inflectional morphemes to words. This does not quite cover Sapir's concepts of type III (concrete relational), as will be shown. However, a language which has these concepts and so belongs to Sapir's mixed-relational types C and D necessarily has a fairly high value for the inflectional index; the inverse relation does not necessarily hold.

The fourth parameter refers to a topic discussed by Sapir as important for the morphological structure of a language but not included in his final formulation. This is the order of subordinate elements in relation to the root. The main distinction here is between the use of prefixes and suffixes. The prefixial index P/W, is the ratio of prefixes to the number of words, and the suffixial, S/W, is the ratio of suffixes to the number of words. An index of infixing, that is, of the number of subordinate elements which are incorporated within the root could likewise have been calculated, but the occurrence of infixes was so rare in the particular languages investigated that it seemed justifiable to omit it. There are an indefinite number of further types of relative position of subordinate elements to the root, for example, containment, as in the Arabic imperfective prefix of the second person feminine which surrounds the verb morpheme, e.g. *taqtuli·*, "thou (fem.) killest," in which the second person feminine morpheme is *ta----i·* while "kill" is *-q-t-l* and

"imperfect tense" is *-u-*. There is likewise intercalation, again found in Semitic, in which a portion of the subordinate element precedes or follows the root and another portion is incorporated. All these devices are so rare that, at least for the languages treated, it did not seem worth while to calculate indices. Sapir's symbolism, which he considers a kind of technique along with isolation, agglutination, and fusion, really belongs here. Sapir's symbolism or internal change is, in my opinion, simply the infixing of an inflective element, e.g., the preterite *-a-* in *sang*. When the elements are derivational, as in Indonesian languages, the process is usually called infixing. This brings out the fact that there are two distinct considerations involved in Sapir's use of the term symbolism, position and regularity. An infixing process might well be regular, in which case the construction would be agglutinative. In fact, though, this hardly ever occurs.

The final parameter has to do with the devices employed for relating words to each other. It therefore brings in syntactical as well as morphological considerations. There are three devices that languages may use, inflectional morphemes without concord, significant order, or concord (agreement).

Languages using the first two devices belong to Sapir's pure-relational category, while those using concord are mixed-relational. The inflectional index discussed above would include both nonconcordial and concordial inflectional morphemes. This index, which may be called the index of gross inflection, is therefore of limited use for the present problem. It might seem that by distinguishing between concordial and nonconcordial inflectional morphemes and by assigning the words without inflectional morphemes to the isolating class, it would be possible to make a clear threefold division. The extent of isolating, inflectional, and concordial techniques could then be calculated by three indices based on the ratio of each of these types to the total number of words. There are a number of complications that prevent such a simple procedure. Many languages, for example, Latin, merge concordial and nonconcordial features in the same inflectional morpheme. Thus, the *-um* of the Latin masculine accusative singular of adjectives has two concordial features, gender and number, and a pure inflectional one, case. In such cases, the procedure adopted is to count the same morpheme a number of times, one for each distinct feature. Another difficulty arises in regard to order. Order probably always has some value in relating elements even where inflection exists. We relate the accusative to the nearest verb even where the order is not fixed. Order may even be fixed although other means are present to indicate which words are in construction. This is largely true of German, for example. Significant order should perhaps be restricted to cases where a change of order produces a change of constructional meaning. The criterion employed was closest to this latter one, but easier of application. The absence of an inflectional morpheme in a word was taken as an indication that the method of relating it was order. If we call each instance of the use of a principle to indicate relations between words in the sentence a nexus, then three indices, O/N, Pi/N, and Co/N were calculated, where O = order, Pi = pure inflection, Co = concord, and N = nexus.

The following, in summary, are the typological indices which have been described:

(1) M/W = index of synthesis
(2) A/J = index of agglutination
(3) R/W = compounding index
(4) D/W = derivational index
(5) I/W = gross inflectional index
(6) P/W = prefixial index
(7) S/W = suffixial index
(8) O/N = isolational index
(9) Pi/N = pure inflectional index
(10) Co/N = concordial index

The validity of these indices assumes that we can define the units employed consistently and in such a manner that they may be applied to all languages. In fact, there is hardly one of the units employed in the above formulas which does not admit of a number of alternative definitions. The choices made here are dependent on the particular purposes of the study. We always ask what it is that we want to measure. In certain cases there seems to be no good reason for the choice of one alternative over the other from this point of view, and a purely arbitrary choice was made since some basis of decision had to be reached. It may be of some comfort to note that the theoretically wide range of choice of definitions for certain units only bore on decisions for a relatively small proportion of difficult instances. As evidence of this, the results of the indices calculated for a passage of 100 words of English in 1951, and arrived at by methods no longer fully recoverable by introspection, may be compared with indices for a 100-word passage done recently in accordance with the methods outlined here.

	1951	1953
Synthesis	1.62	1.68
Agglutination	.31	.30
Compounding	1.03	1.00
Prefixing	1.00	1.04
Suffixing	.50	.64
Gross inflection	.64	.53

It should be emphasized that other alternatives than those chosen here for the definition of units are equally possible and probably preferable for certain other purposes, for example, writing a grammar of a language.

In the following section the chief problems encountered in defining the units employed in the indices are discussed. These refer to the morph, the morpheme, agglutinative constructions, the distinction of root, derivational and inflexional morphemes, and the word. Nothing approaching an exhaustive treatment of the problems is attempted here. The purpose of the present discussion is merely to point out the chief problems encountered in this study and the reasons for the particular solutions which were adopted.

Basic to the synthetic index as well as most of the others is the possibility of segmenting any utterance in a language into a definite number of meaningful sequences which cannot be subject to further division. Such a unit is called a morph. There are clearly divisions which are completely justified and which every analyst would make. For example, everyone would divide English *eating* into *eat-ing* and say that there were two units. There are other divisions which are just as clearly unjustified. For example, the analysis of chair into *ch-*, "wooden object," and *-air*, "something to sit on," would be universally rejected. There is, however, an intermediate area of uncertainty in which opinions differ. Should, for example, English *deceive* be analyzed into *de-* and *-ceive*? It is this intermediate area with which we must be able to deal. We start with a set of forms that will be hereafter called a square. A square exists when there are four meaningful sequences in a language which take the form AC, BC, AD, BD. An example is the English *eating* : *sleeping* :: *eats*:*sleeps*, where A is *eat-*, B is *sleep-*, C is *-ing* and D is *-s*.[2] Where a square exists with corresponding variation of meaning, we are justified in segmenting each of the sequences of which it is composed. Once it has been segmented, each of its segments may then be tested to discover if it also is a member of a square. If it is, it in turn will be segmented into two morphs. If it is impossible, then we have reached the limit of analysis and cannot divide further. A test of correspondence of meaning is

[2] One of the four elements may be zero provided the sequences in which it occurs are free forms, i.e., may occur in isolation. For example, hand:hands :: table:tables is a valid square, in which A is *hand*, B is *table*, C is *zero*, and D is *-s*.

applied to avoid such squares as *hammer*:*ham* :: *badger*:*badge*. A square conforming to these conditions described will always give us valid, generally acceptable analyses. It is too severe, however, in that it excludes some segmentations which everyone would want to accept. The first extension which we make is the following. A sequence which occurs with a member of a square is also recognized elsewhere if with regard to this member (a) the sequence of phonemes is identical except for automatic changes (for which see below) and (b) if the meaning is the same. On this basis we recognize a segmentation of *huckleberry* into *huckle* plus *berry*, since *berry* itself is a morph elsewhere. This leaves *huckle-* also as a morph although it never occurs in a square. If now *huckle-* were to occur in some other combination, we would recognize a segmentation there also and so add a new morph. This process is continued until we reach a sequence that does not recur in any other combination. In this instance we have reached it with *huckle-*.

A further extension must be made in the case of what might be called a formally defective square. We should like to analyze men into two morphs, one with the meaning "man" and the other "plural," but there is no square into which it can be put. For example, *man*:*men* :: *boy*:*boys* is formally defective. We set up the following rule. If there can be found a square like the one just cited in which *boy*:*boys* itself is a pair in another valid square, e.g., *boy*:*boys* :: *lad*:*lads*, and if *man* may always substitute for *boy* and *men* for *boys* and produce a grammatical (even though at times semantically improbable) sentence, then *man*:*men* may be subjected to a segmentation analogous to that of *boy*:*boys*, and *men* may be considered two morphs. In the case of *sheep*:*sheep* :: *goat*:*goats* we recognize two morphs for the plural of sheep, one of which is a zero. Such analyses are not to be confused with segmentations into two or more semantic categories, where no valid square exists for substitution. In Latin, for example, we cannot analyze *-us* nominative singular into two morphemes, nominative and singular. The square *-us*:*ō*::*ī*:*īs*—nominative singular:dative singular::nominative plural:dative plural—does not have a pair which can be substituted for members of a formally perfect square, and hence the segmentation of these forms is not permitted. Corresponding to formally defective squares we have those which are semantically defective. Here, if there are parallel non-automatic variations, the analysis is permitted even though definite meanings cannot be assigned to the morphs. Thus, the sets *deceive*:*receive*::*decep-tion*:*recep-tion*::*decei-t*:*recei(p)t* justify the segmentations *de* + *ceive* and *re* + *ceive*. This rule permits the usual assumption of morphs for derived forms of the verb in Semitic. In view of the variety of meanings in instances of this kind, it would be difficult to operate without it.

There are certain extensions of the morph concept which are rejected here as inconsistent with the purposes of this study, although entirely legitimate for other purposes. No discontinuous morphs which contain segments in two different words are accepted. This is understandable, since we wish to measure the ratio of morphemes to words and we therefore want each word to contain a definite number of morphemes restricted to the word itself. We likewise do not include meaningful units simultaneous with grammatical stretches longer than the word, e.g., intonation patterns of the sentence. The reason again is clear. We wish morphemes to be parts of words and this they cannot be if they are simultaneous with an entire sequence of words. In this connection it should be noted that neither the index of synthesis nor any of the others utilized in the present study is a general measure of total complexity of a language. Intonational patterns and certain other items which contribute to the over-all complication of a language are not included.

The next step to be taken after the identification of morphs is the establishment of more complex units, morphemes, with morphs as their members. It is this aspect of the problem which has been most frequently discussed in the literature as the chief subject-matter of morphemic analysis by Harris, Hockett, Bloch, Nida, and others. In general, the principles set forth by Nida are adequate (1948, pp. 414–441). These involve the generally recognized criteria of sameness of meaning (here applied strictly) and of complementary distribution and the requirement that if morphs of varying phonemic shape are to be assigned to the same morphemic unit, there must be at least one nonvarying unit with at least as wide a distribution.

On this point, however, for reasons to be explained, it is inadvisable to accept the complementations allowed by Nida in accordance with his rule that "complementary distribution in tactically different environments constitutes a basis for combining different forms into one morpheme only on the following condition: that some other morpheme—belonging to the same distribution class, and having either a single phonemic shape or phonologically defined alternant shapes—occurs in all the tactically different environments where the forms in question are found" (p. 421). For example, in Arabic there are pronominal suffixes indicating possession when added to the noun and another set indicating the verb object when added to the verb. These are tactically different: i.e., the verb in general could not be substituted for the noun or vice versa. The existence of *-ka*, "second person singular masculine," and other phonemically identical forms in both series would, on the basis of Nida's rule, allow us to unite the morphs of the first person singular: *-ī* and *-ǧa*, noun possessive, and *-nī*, verb object, as constituent morphs of the same morpheme. This alternation is, of course, irregular and in reckoning our agglutinative index, if we accepted this alternative, we should have to call any construction involving one of the first person singular suffix forms irregular, or nonagglutinative. We should thus be penalizing Arabic, as it were, for having some degree of regularity in these forms. In a language with two totally different sets of pronouns in these uses, Nida's rule would not allow complementation; hence there would be no irregular alternations from this source, although, on a common-sense basis, we should call this the more irregular situation. Hence only members of the same structural set, i.e., those substitutable for each other in the same tactical environment, are considered as possible alternants of the same morpheme.

Given certain morphs as alternants of the same basic morpheme units, we can define an agglutinative construction. It seems constant with the traditional use of the term to consider its basic reference to be to morphological regularity. However, the term regular has been used in a number of different ways. In the discussions of Bloomfield, Wells, and others it is now usual to distinguish types and degrees of regularity and irregularity. The definition of *regularity* adopted here is that which appears to be closest to actual usage in typological discussion. This is the requirement that all the varying phonemic shapes of the morph be derivable from a nonfictive (i.e., actually occurring) base form by rules of combination which hold for all similar combinations throughout the language. This is usually called automatic alternation. The case of a morpheme which does not have alternating morphs, i.e., which involves the same phonemic sequence in all occurrences, is a limiting instance which is, of course, also reckoned as automatic. Sometimes one choice of a base form gives automaticity, i.e., predictability from base forms, while another does not. In cases of doubt those base forms are chosen which give the greatest degree of automaticity for the total description of the language. This is admittedly not

a very precise rule, but in practice this does not prove to be a major difficulty.

We define automaticity as the property of the entire morpheme where every morph is in automatic alternation with every other. The morphs may often be grouped in subalternating sets. It is not sufficient for these to be in automatic alternation. The English plural morpheme has the statistically most frequent subset *-s -z -ez*, which are in automatic alternation. However, there are other alternants, e.g., *-en*, *-zero*, etc., which are not in automatic alternation with *-s -z -ez* as a whole. Hence the English plural morpheme is not automatic.

The possibility of calculating compounding, derivational, and inflectional indices depends on our ability to distinguish root, derivational, and inflectional morphemes. Of these the root class is probably the most difficult of formal definition, but the easiest of recognition. By this is meant that in actual practice, there is virtually complete agreement as to which morphemes are to be regarded as root morphemes. The root position in the word is characterized by a large and easily extendible membership and concreteness of meanings. The contrast with inflectional morphemes is greatest in these respects, since these latter tend to be few in number and relational as well as abstract in meaning. Another characteristic would generally be agreed on. Every word must have at least one root morpheme. Hence in a one-morpheme word, that morpheme is necessarily a root. In contrast, derivational and inflectional morphemes need not occur, and there are some languages, so-called root or isolating languages, in which derivational and inflectional morphemes are rare or perhaps do not occur at all. Derivational morphemes may be defined as morphemes which, when in construction with a root morpheme, establish a sequence which may always be substituted for some particular class of single morpheme in all instances without producing a change in the construction. If the class of single morphemes for which the derivational sequence may substitute contains one of the morphemes in the derivational sequence itself, we call the sequence endocentric; if not, then it is exocentric.

For example, *duckling* in English is a derivational sequence, since it may be substituted anywhere for *goose*, *turkey*, etc. without change of constructional meaning. Since *duck* is included in this class of single morphemes for which *duckling* may substitute, *-ling* is here an endocentric derivational morpheme. *Singer* is an exocentric sequence, since the class of single morpheme sequences for which *singer* may substitute consists of single-morpheme nouns only, and does not include the verb *sing*. Hence *-er* is an exocentric derivational morpheme.

We can now define the inflectional morpheme simply as a nonroot, nonderivational morpheme making the three classes exhaustive and mutually exclusive. Inflectional morphemes, like derivational, need not occur at all in any particular language. When it is part of a word pattern, however, its appearance in the appropriate position is compulsory like that of the root. One member of the class is frequently zero. In these instances, the absence of an overt phonemic sequence shows itself as significant because the word in this form has definite syntactic limitations on its uses; e.g., the nominative singular in Turkish or the noun singular in English.

We finally reach what is in some ways the most difficult problem, the definition of the word unit. It is clearly fundamental to the purpose of the present study inasmuch as all the indices involve the number of words. In most instances this is explicit; sometimes it is tacit, as in the index of agglutination, in which the number of morpheme junctures is always one less than the number of words. There is at present no general agreement on this topic. Some deny the validity of the word as a linguistic unit. Others admit it, but deny that it need be taken into account in the description of a

particular language. Some say the word is definable only for each language in a separate *ad hoc* fashion. Some define it in phonological, others in morphological terms. In practice, however, the word continues to be the key unit of most actual language descriptions. Of the two basic types of overall definitions of the word, the phonological and the morphological, the former is clearly insufficient for the purposes of the present study. In phonological definitions, we define the word in terms of some single phonological characteristic, or through a combination of characteristics which serve as markers. These markers are usually stresses or boundary modifications of phonemes; i.e., junctures. Besides the fact that the use of phonological markers to define the word sometimes leads to the isolation of individual units which we should not wish to call words on other grounds, many languages do not have such phenomena, so that a phonological procedure cannot lead to a universal definition. The other basis has been called morphological, since it is based on the distribution of meaningful elements. Of definitions of this kind, Bloomfield's characterization of the word as the minimum free form is the most satisfactory in that it is universally applicable and points in the right direction to freedom or absence of freedom, that is, bondage as the basic criterion. The actual test of freedom, ability to occur in isolation, is, however, difficult to apply in practice and leads to unusual results. For example, *the* in English would not be a word by Bloomfield's test.

The procedure adopted here can only be briefly outlined. It has led to results satisfactory for this study in the relatively few doubtful cases regarding the existence or nonexistence of a word boundary in the languages under consideration. Instead of asking whether a particular minimal form is bound or free in general, as is usually done, the present treatment is in terms of morphs in particular contexts. This allows us, for example, in Latin to make *ab*, "from," a free form as a preposition but a bound form as a verb prefix in *abduco*, "I lead away." What is specified as bound or free is not a morph as such but a contextually determined class of mutually substitutable morphs. Such a class is here called a morph substitution class (MSC). This notion is expanded to include a sequence of morpheme substitution classes which may in all circumstances be substituted for a particular MSC and none of whose members are identical in membership with it.[3] It is convenient to use the term *nucleus* to cover both individual MSC's and such substitutable sequences. Having broken up an utterance into nuclei in this manner, we now test each nucleus boundary to see if it is a word boundary or not. A nucleus boundary is a word boundary if it is possible to insert an indefinitely long sequence of nuclei. If it is an intraword boundary, either no nucleus can be inserted or a fixed maximum number can be. For example, in the sentence "the farmer killed the ugly duckling" there are nine morphemes: (1) the (2) farm (3) er (4) kill (5) ed (6) the (7) ugly (8) duck (9) ling; seven nuclei: (1) the (2) farmer (3) kill (4) ed (5) the (6) ugly (7) duckling; and six words: (1) the (2) farmer (3) killed (4) the (5) ugly (6) duckling. There is an intraword boundary at "kill-ed" because no nucleus may be inserted.

[3] This is necessary in order to exclude endocentric phrases in which a sequence of words can always be substituted for the head or chief number. A sequence of adjectives followed by a single-morph noun would be a nucleus were it not for the proviso of dependence among its members. Adjectives are not bound to nouns in English, for example, because they occur in predicative adjective constructions also. For the basic ideas of the MSC and the derivational sequence, I am largely indebted to the stimulus of the writings of Zellig S. Harris and Rulon S. Wells. The resemblance to the notions of focus class and expansion of the latter writer is particularly close. See especially R. Wells, 1947, pp. 81–117.

TABLE 1

	Sanskrit	Anglo-Saxon	Persian	English	Yakut	Swahili	Annamite	Eskimo
Synthesis	2.59	2.12	1.52	1.68	2.17	2.55	1.06	3.72
Agglutination	.09	.11	.34	.30	.51	.67	...	.03
Compounding	1.13	1.00	1.03	1.00	1.02	1.00	1.07	1.00
Derivation	.62	.20	.10	.15	.35	.07	.00	1.25
Gross inflection	.84	.90	.39	.53	.82	.80	.00	1.75
Prefixing	.16	.06	.01	.04	.00	1.16	.00	.00
Suffixing	1.18	1.03	.49	.64	1.15	.41	.00	2.72
Isolation	.16	.15	.52	.75	.29	.40	1.00	.02
Pure inflection	.46	.47	.29	.14	.59	.19	.00	.46
Concord	.38	.38	.19	.11	.12	.41	.00	.38

On the other hand at the boundary between "farmer" and "killed" there is no fixed maximum number of insertable nuclei. We may talk of the "farmer who killed the man who killed the man who . . . killed the ugly duckling." The contradiction with the phonological word is in certain cases merely apparent. Thus in Latin the enclitic *-que*, "and," which is reckoned as a syllable with any preceding sequence in locating the stress which serves as a phonological word marker, is also part of the word by the present test. *Dō'minus*, "the lord," and *dō'minús* in *dō'minúsque*, "and the lord," are not members of the same MSC because they are not substitutable for each other. *Dō'minús-* belongs to the same nucleus as *legātús-*, *puér*, and this class is dependent on the class of the following *-que*, *-ve* since it must be followed by it and hence belongs to the same word. Even with monosyllables, where there is no stress shift, *mū's*, "mouse," and the *mūs* of *mū's-que*, "and the mouse," are members of different nuclei, since the former can be substituted only by *dō'minus*, *puér* etc. The latter only by the class of *dō'minús-*, *puér-*.

Table 1 shows the calculated indices. The languages selected are chiefly those frequently cited as examples of specific types in the existing literature in typology, subject to the limitations of my own knowledge of specific languages. Instead of Turkish, the related Yakut was selected as an example of an agglutinating language, since the extensive Arabic borrowings in Osmanli Turkish have led to irregularities in the vowel harmony and in other respects to such an extent as to render it untypical. Two ancient Indo-European languages, Anglo-Saxon and Sanskrit, were chosen, and two modern languages of the same German and Indo-Iranian branches, modern English and Persian, were also selected to illustrate long-term change in type. Annamite was selected as a representative root isolating language, Eskimo as polysynthetic, and Swahili as an agglutinative, concordial Bantu language.[4] (See accompanying table.)

On the basis of counts such as these, the next step would be to define terms like analytic, synthetic, agglutinative, and pre-

[4] The passages selected were 100 words long for each language, as follows: Sanskrit, *Hitopadésa*, ed. Max Mueller, p. 5, *varam ekas*, ff.; Anglo-Saxon, *An Anglo-Saxon Reader*, J. W. Bright (New York, 1917), p. 5, *hit gelamp gio*, ff.; Persian, Pizzi, I., *Chrestomathie Persane* (Turin, 1889), p. 107, *ruzi Ibrahimi*, ff.; English, New Yorker, December 13, 1952, p. 29, *Anyone who*, ff.; Yakut, *Ueber die Sprache der Yakuten* (St. Petersburg, 1851), p. 29, *min bāyāsä* ff.; Swahili, C. Sacleux *Grammaire Swahilie* (Paris, 1909), p. 321, *Kiyana mmoja*, ff.; Annamite, M. B. Emeneau, *Studies in Vietnamese* (*Annamese*) *Grammar* (Berkeley and Los Angeles, 1951), p. 226, *mot hom*, etc.; Eskimo, W. Thalbitzer in *Handbook of American Indian Languages*, Part I, ed. Franz Boas (Washington, 1911), p. 1066, *kaasasurujuyuaq*, ff (phonemicized and slightly normalized to conform to Kleinschmidt's grammatical description).

fixing after plotting a frequency distribution curve. There are too few languages here to make this procedure feasible. However, even cursory inspection of the indices set forth here shows that, if we define an analytic language as one with a synthetic index of 1.00–1.99, synthetic as 2.00–2.99, and polysynthetic as 3.00 +, the results would conform to the usual nonquantitative judgments. Similarly we might call a language agglutinative if its agglutination index is over 0.50 and similarly for the others.

The present results should also be confirmed by further counts, since they were obtained only for single passages of 100 words each, besides which the probable error should be calculated for each. One might well suspect differences based on the style of the passages selected. The synthetic index was calculated in English and German for a number of passages of varying style, however, with remarkably concordant results.

English

Ladies' Home Journal, January 1950, p. 55	1.62
R. Linton, *Study of Man*, p. 271	1.65
O. J. Kaplan, *Mental Disorders in Later Life*, p. 373	1.60

German

Baumann, *Nama Folk-tale*	1.90
Ratsel, *Anthropogeographie*, p. 447.	1.92
Cassirer, *Philosophie der Symbolischen Formen*, p. 1	2.11

This, of course, is no substitute for statistical evaluation. Another topic that can be studied by the present method is the general direction of historical changes in language over an extended period. The coincidences between Sanskrit and Anglo-Saxon on the one hand, and Persian and English on the other hand, are striking. The direction of change for virtually every index is the same from the older to the more recent language. Were more conservative Indo-European languages such as the Slavonic chosen, the results might well be different.

The present study is to be evaluated as purely a preliminary attempt. Some indices may well have to be eliminated and others substituted. The specific definitional choices may likewise be subject to alteration in future studies. However, the general method of the calculation of indices based on text ratios of carefully defined linguistic elements has, I believe, definite value for typological studies.

FIRST THOUGHTS ON SYNTACTIC INDICES

FRED W. HOUSEHOLDER, JR.
INDIANA UNIVERSITY

0. Introduction
1. Particle-word ratios
2. Before-After ratio
3. Movability index
4. Depth, transformations
5. Style vs. language

0. Greenberg's use of objective measures derived from morphology (for the most part) as a device for typological classification and John Carroll's use of similar measures in a stylistic study of English[1] suggest that some sort of useful numerical indices might be obtained from syntactical characteristics of languages.

Some of the most obvious and frequently mentioned syntactical differences do not easily lend themselves to this technique; for instance, the presence or absence of something which can reasonably be called a 'case system', or 'tense', or 'aspect', the use of structural signs for the subject of intransitive verbs similar to those for either the object or else the subject of transitive verbs, the use of the orders SOV, SVO, VSO, VOS, OSV, or OVS as necessary or preferred in transitive sentences, the presence or absence of an inclusive-exclusive contrast in person markers, of a dual number (or of number as a category at all), of suppletion in the verb system—all these, and many more like them, are difficult to reduce to a meaningful number.

1. But there are others. Greenberg himself uses one set (his battery of three based on the relative use of three methods of marking a nexus,—what he calls 'Isolation'—apparently including word-order and particles as markers—'Pure inflection' and 'Concord'—i.e. inflection which also serves some other function) which is largely or wholly syntactical, and some of his others suggest easy extensions to syntax. His measures of synthesis (ratio of morphs to words) and compounding (stems to words) suggest a similar measure, either of words per sentence (used by Carroll as a stylistic measure, and probably unreliable as a typological one) or of words per phrase (assuming we can get a workable and comparable criterion for identifying phrases). His 'agglutination' index could be matched by a 'sentence sandhi' index; but again (a) stylistic variability would make it unreliable, and (b) the information required is often difficult to get. His derivation index (derivative affixes per word), gross inflection (inflectional affixes per word), prefixing (prefixes per word) and suffixing (suffixes per word) suggest similar ratios involving particles (again assuming a consistent and comparable definition): ratio of particles to words (this seems less liable to stylistic variation than particles to phrases or sentences), of prepositive particles (proclitics, prepositions, many conjunctions, etc.) to words, and of postpositive particles (enclitics, postpositions, emphatic particles, sentence-final particles, etc.) to words. I have worked these figures out for a few simple 100-word passages of English and of classical Greek.[2]

	Eng. I	II	III	Plato	Thuc.	Xen.
Pre/w	.31	.31	.38	.15	.20	.15
Post/w	.01	.01	.00	.17	.08	.11
Total	.32	.32	.38	.32	.28	.26

[1] John B. Carroll, "A Factor Analysis of Literary Style" in Style in Language, Part IX, ed. T. A. Sebeok (New York and London: The Technology Press of M. I. T., and John Wiley and Sons, Inc., 1960).

[2] In English, one passage in Greenberg's article beginning on page 195, one from *TIME* magazine, April 4, 1960, page 20 (beginning with "Amidst") and another from page 55 ("The acid bath"); in Greek, one from Plato's *Apology* (beginning at 21A), one from Thucydides (2.7. 1–2), and one from Xenophon (*Anabasis*, beginning at the beginning).

These ratios seem to be fairly stable, though there is plainly some stylistic variation. In regard to the total particle ratio, English and Greek do not differ greatly (though other languages—e.g. Turkish—would), but there is a striking difference in the proportion of postposed vs. preposed particles, where Greek shows many more postposed particles than English (in one sample, Plato, these are even more numerous than the preposed ones). The ratio of words to sentences, mentioned above, shows the following values for the same 6 samples.

	E I	E II	E III	Pl.	Th.	Xe.
W/S	16.7	20	12.5	7.7	33.3	14.3

Here the variation in the Greek samples is greater than that between English and Greek; obviously this measure is useless for typology (at least in this form), however good it may be for stylistics.

2. Another measure which might be compared to the prefixing-suffixing ratios is one concerned with the number of words in constructions which precede their heads, as compared with the number which follow. This requires some arbitrary decisions in the case of so-called exocentric constructions; I have chosen to consider the preposition as head of prepositional phrases (on the ground that "The master is in the house" has more in common with "The master is in" than with "The master is the house"), subordinating conjunctions as heads of their clauses, and verbs or verb-phrases as heads of sentences. The best measure appears to be that of words before heads to words after heads (rather than of these to total words). I give the figures here for three Greek and three English passages.[3]

	Hu.	Gr.	Gi.	Xe I	Xe V	Pl.
B/A	.27	.31	.29	.49	.46	.58

While there is obviously some stylistic range here (in both languages), in this case, at least, the difference between the languages is much greater. It is interesting that the greatest difference is between Hunt and Plato, since both of these passages involve many short, simple, colloquial sentences (in contrast to the other four).

3. A measure of flexibility of order might be a desirable figure to have; the ideal, perhaps, if one could calculate it, would be P/8, possible permutations divided by the number of sentences. A figure which will do almost as well, and is much easier to arrive at, may be had by considering each word in turn and noting how many other positions in the sentence it could be shifted to (if no other changes are made), and dividing this total M by W, the number of words. If we take the Greenberg 100-word passage already used, we get a flexibility ratio of .13, and for the Gibbon passage .15, while for Plato we have .68 and for Xenophon (1.1) .92. If we extend our rules to consider not only single words, but also phrases of two and of three words (but none longer), we may get for Greenberg a ratio of .18, for Gibbon .31, for Plato 1.09 and for Xenophon 1.45. Again the variation within each language seems much less than the difference between the two.

4. Many more measures are probably worth testing, though some of them may prove to be too bound up with style, rather than language. For instance, at the last Linguistic Society meeting V. H. Yngve of M. I. T. presented an interesting procedure for analyzing sentence structure[4] which yields for each sentence an integer called

[3] The English passages are (1) a 378-word stretch from a detective novel by Kyle Hunt, To Kill a Killer (New York: Random House, 1960), p. 44; (2) a passage of equal length from Greenberg's article (including the 100-word passage used before), and (3) a 100-word passage from Gibbon's Decline and Fall of the Roman Empire from the beginning of Chapter 47. The Greek passages are: the first 148 words of Xenophon's *Anabasis*, 195 words from the *Anabasis* 5.1. 8–11, and the same 100 word passage of Plato that was used before.

[4] A Model and an Hypothesis for Language Structure.

'maximum depth'. It is at least possible that the average maximum depth over a sufficiently large random sample of a language would prove to be characteristic for the language. A very similar measure might be based on the average number of transformations required to generate each sentence in such a random sample; this, however, requires that consistent transformational grammars on the same pattern are available for all the languages being compared. Even if these two turn out to be unsuitable for typological purposes, they may still be interesting for stylistic comparisons.

5. Greenberg's own indices do not seem to be entirely style-free (or mutually independent: Kroeber's arrangement in rank orders shows 3b and 4b with exactly the same order, 1 and 3c extremely similar); when I calculate them for the same one-hundred word sequence from his own paper which we have used before, I get 1.64 for M/W, .35 for A/J, 1.05 R/W, .36 D/W, .22 I/W, 1.5 P/W and .54 S/W. This suggests that M/W (which he had already tested with a variety of passages) is indeed fairly constant, but D/W and I/W distinctly less so; either of these figures would substantially alter the rank order of languages. Probably longer stretches than 100 words would help to even things out, but even so derivation and compounding (particularly in Sanskrit, for instance) can vary considerably with style.

So we come to the last of our first thoughts. There are probably at least a half-dozen reasonably consistent syntactical indices which can be used for typological purposes; like Greenberg's morphological indices, all of them (as well as others whose value cannot be decided in advance) must be thoroughly tested for degree of freedom from stylistic influence and for general consistency from language to language.

TYPOLOGY OF DENSITY RANGES I: INTRODUCTION

C. F. VOEGELIN, A. K. RAMANUJAN, F. M. VOEGELIN
INDIANA UNIVERSITY

0. Scope
1.0. Density ranges in morpho-syntax
1.1. Range of stem sequences in text and inventory
1.2. Range of affixes and operators in text and inventory
2. Allomorphic density range
3.1. Phonemic density range in text and inventory
3.2. Implication of phonological and other subsystem typology
4. Non-binary nature of density ranges

0. A brief three-stage history of the whole language typology problem may set it in perspective. Underlying its manifest 'history', work in typology reflects broad 'interests' or 'needs' which are not satisfied in structural linguistics, with its emphasis on what is unique in a given language. All typologies—subsystem as well as whole language—look for what is comparable among natural languages, whether related or not.

0.1. The 19th century linguists included ethnocentric considerations in their typology. On the linguistic side, their typology was set up under the assumption that some morphosyntactic features of a given language will serve as an index or indication of the type of that language as a whole. Such typology attempts to measure the whole language—by implying a correlation between parts of the language treated in the typology and parts not treated.

0.2. Sapir excluded all extra-linguistic considerations in his typology. He eschewed the evolutionary and ranking interests of his predecessors, but otherwise followed the main logic of whole language typology—a logic of metonymy in which a few parts (i.e. a low number of morphosyntactic features) represent the whole. Like the 19th century typologists, Sapir cited features for a given language as though the language in question were fully described. From the inventory of any of its features—in effect, a vertical storage of all relevant features—Sapir would select certain ones for purposes of typologizing. The dozen terms that Sapir used for his features or indices are not equally coordinate as such, but are coordinate in terms of three axes or bases: formal processes (*isolating*, *symbolic*, *affixing* by *prefixes* or *suffixes* or *infixes*); semantically oriented word construction (*analytic*, *synthetic*, *polysynthetic*); semantically oriented morphophonemics (*fusing-inflective*, *juxaposing-agglutinative*).

Sapir typologized exclusively from a vertical storage source of information for a given language ('lists and rules'), thereby imposing a unidirectional procedure: it became necessary to structuralize before typologizing a given language—a necessity that must have seemed so obvious as to be not worth mentioning. But for other considerations, see **3.2**, below.

0.3. Greenberg did more than quantify Sapir. He changed the basis of typology from a vertical storage base of information, as used by Sapir, to a horizontal or syntagmatic text frequency count. In programming a translation computer, one is faced with much the same kind of problem that Greenberg proposes for typologizing. In either case, it is possible to program a computer or typologize a given language with one set of texts.

0.4. The three-stage history of whole language typology (**0.1–0.3**, above) is so isolated from the history of subsystem typology, that workers in one development do not in general appreciate work in the

other.[1] Whenever new entities become isolable in structural linguistics, they lend themselves to typologizing—from the phonological typologies of Trubetzkoy to the one suggested by Harris for kernel sentences in transformational grammar.

We now attempt to go beyond phonology, without assuming that we have, in some new or old way, hit upon a short-cut to characterizing a whole language through a few salient features. When we state the phonemic density ranges in languages *A . . . n* it becomes possible to compare the subsystem in question among languages *A . . . n*. We move cautiously from one part of a language typologized to another part. If the typology of one part implicates that of another part, that would be important—but discoverable only after preliminary typologizing of more than one system across languages.

0.5. For each part of a language being typologized, we have recourse to (1) a text, as well as to (2) 'inventory'. By 'inventory' we mean: the vertical storage as given in a grammar of a more or less fully described language. But we wish to typologize languages for which full length grammars are not yet available but which have been sufficiently analyzed to permit recognition of the overall structure. For such languages, still under analysis, we use 'inventory' to refer to some approximation of a grammar, which at a minimum would contain numerous forms obtained by ancillary eliciting from the text cited. Here, (1) text and (2) inventory are not taken as independent variables; rather, (1) is related to (2) for each density range considered.

0.6. With the exception of the term used in the title, 'the typology of density ranges' uses only those terms already defined for structural linguistics.

1.0. Statements in this section are concerned with various morpheme classes. The assumption is that nouns (N) in one language are typologically comparable to nouns (N) another language only in the context of the structural basis of each language. Thus, since the total number of stem classes are three for Hopi, for example, but five for Kannaḍa, a translation between these languages would show that some morphemes which belong to verb class in Hopi would translate into Kannaḍa verb class, while others would translate into Kannaḍa adjective class. So also Kannaḍa morphemes in adverb class would translate either in Hopi verb class or Hopi particle class, and so on. This rather general caveat incidentally points up the relevance of typology to translation problems.

1.1. In compounding, the minimum number of stems is two, though some languages permit more than two stems in one word. Where there is only one stem in a word, there is no compounding. For range of stem sequence in words, however, compounding, as usually identified, is merely one magnitude in the range which actually extends from zero (as when prefix-suffix sequences without a mediating stem are counted as words), to one, to more than one stem (thereby compounding). There exist some languages in which the range of stem density is nil be-

[1] For whole language typology, see Kroeber's paper and Greenberg's paper in this issue of IJAL, and the bibliographic references by them. For bibliography of subsystem typology, beginning with Trubetzkoy, see Charles F. Hockett, Manual of Phonology (IUPAL 11 = IJAL 21 No. 4, 1955), and subsequently C. F. Voegelin, On Developing New Typologies and Revising Old Ones, SJA 11.355–60 (1955); Six Statements for a Phonemic Inventory, IJAL 23.78–84 (1956); Subsystems Within Systems in Cultural and Linguistic Typologies *in* For Roman Jakobson, The Hague (1956); The Scope of Whole System ('Distinctive Feature') and Subsystem Typologies, (with John Yegerlehner) Word 12.445–54 (1956); Linear Phonemes and Additive Components [SGC], Word 12.429–43 (1956); and with reference to Harris, review of Syntactic Structures by Noam Chomsky, IJAL 24.229–31 (1958); John Yegerlehner and Florence M. Voegelin, Frequencies and Inventories of Phonemes from Nine Languages, IJAL 23.85–93 (1957); Henning Spang-Hanssen, Probability and Structural Classification in Language Description (Copenhagen, 1959).

cause the minimum number of stems and the maximum number of stems is exactly the same. The lower limit or minimum density range for some languages is one stem rather than zero stem—that is, for languages which have all prefix-suffix occurrences mediated by a stem. The upper limit of maximum density range for some languages is one stem rather than more than one, but for other languages is more than one—up to a half dozen or even more.

The minimax gives the range rather than the average of stem density; and it gives the range for each particular morpheme class of stem as verb (V), noun (N), particle (P), and so on, rather than for stems as an undifferentiated overall class, hereinafter called the major morpheme class.

We want the differentials selected for the typology to be sensitive to the differentials that obtain in the languages to be typologized. If, now, we were to state the minimax for stem density in terms of an overall major morpheme class, we would obtain a less sensitive differential than if we were to give, as we do, the stem density ranges for each class of stems as N and V and P. For example, in a given language the stem density descends from a maximum of three noun stems (NNN) to a minimum of one (N) while the verb stem density range descends from two verb stems (VV) to one (V). If, now, N and V are merged as a single major morpheme class, the report gives a single stem density range (one to three), and thereby conceals the differential between N and V density ranges. We want to obtain density ranges which will reveal such differentials.

For certain languages, however, the stem density range as given for the text is comparable to the stem density range from the inventory only when recursive extensions of compounding (possible in the inventory) are excluded. We do not, however, exclude mention of the recursive possibility of such compounding, since recursiveness is relevant to maximum density, but rather include such mention in terms of a formula beside the stem density range of non-recursive examples found in the inventory.

By comparing the density ranges of stems in text and inventory, any atypicality in a particular text becomes observable. In this sense any particular text might be unrepresentative of the range of possibilities of a language. This justifies the matching between inventory and text. The text is also useful, for it serves as a sample of a kind of functional yield not obtainable from the inventory alone.

For the text we give the sum of each stem density for each stem class. Thus, the first two sections of a Hopi text yields the following scores for number of noun stems (N), verb stems (V) and particle stems (P): 27 words with single N, 25 words with single P and 7 words with single V. Besides words with single stems, there are three words which include P and V in compound. One may experiment, and make parallel counts for the remaining sections of the Hopi text. Continuing such counting to the end of this text and then beyond to other texts, one could arrive at statistically reliable averages. But we do not necessarily attempt this. We do attempt to count at least enough to raise questions. Having raised such questions as the density range and the relative frequency of different stem classes (single N, P, V—in that descending order; and P-V), we then turn to the inventory in Hopi.

It does not seem to matter how extensive the text is for this preliminary count. In contrast to the first two sections of the Hopi text, the experimental Kannaḍa text is three times as long, and yields the following scores for noun stems (N), verb stems (V), adjective stems (A) and adverb stems (Adv.): 80 words with single N, 33 words with single V, 26 words with single A, and 13 words with single Adv. Besides words with single stems, there are ten words which include or occur with NV in compound, five with NN and five also with VV, and two words for AN. Besides words with two stem compounds, there are also words with three stem compounds: VVV in two words and NNN in

one. Having raised the question of the descending order of stems from information given by the Kannaḍa text sample (N, V, A, Adv., NV; NN in the same rank as VV; so also AN, VVV in the same rank; and finally NNN), we turn to the inventory in Kannaḍa.

Faced with the inventories of Hopi and Kannaḍa, we ask a series of questions, which are given in the following paragraphs, together with general answers. This shows the related uses of the sample text and of the inventory in respect to stem density ranges. The text orients us, or, in other words, suggests questions. The inventory answers these questions, if it is a sufficiently full inventory; without the text, the full inventory turns out to be so rich in its storage that it leads us away from the obvious kind of questions that are raised recurrently in texts:

Is the inventory minimum the same as the text for stem density ranges? The answer is yes for Hopi, and yes also for Kannaḍa.

Have some instances of every word class appeared in the sample text? The answer is yes for Hopi, but no for Kannaḍa. That is, the Kannaḍa inventory includes the stem class of particles, but the sample text does not.

Has some one or more instance of every variety of compounding appeared in the sample texts? The answer is no for Hopi and no also for Kannaḍa.

This brings us to the point of stating minimax in stem density ranges. This is easy to state from the sample texts. It is also easy to state the minimum in stem density range from the inventory. However, the inventory does not give us a correspondingly easy basis for stating the maximum in stem density range. We do consider some of the difficulties for each of the languages sampled, in subsequent installments of this paper, somewhat in the spirit of offering a typology of difficulties.

1.2. Parallel to the cover-term 'major morpheme class' for stem classes, we use the cover term 'minor morpheme class' for affixes and operators. In general, major morphemes have a higher list count than minor morphemes, while minor morphemes have a greater text frequency than major morphemes.[2] Morphemes from both of these overarching classes are typologized with reference to particular stem classes. For example, in the case of affixes we do more than distinguish between prefixes and infixes and suffixes. We distinguish prefixes which appear in sequence with one particular stem class from those which appear in sequence with another class or with either of two particular stem classes—and so on.

Affixes would appear to need no special identification. Though prefixes in general precede stems and suffixes follow them, sequences of prefixes and suffixes appear in some languages without a mediating stem (or alternatively, with zero stem)—notably in Algonquian; and the same affix appears before the stem in some sequences, after the stem in other sequences (as the person marking 'ambifixes' of Hidatsa). In short, the usual notion of an affix as a bound morpheme with unique position class in respect to a stem is not adequate for all types of languages.

We treat operators as coordinate with affixes, and distinguish four kinds of operators: reduplication which is morphemic; phoneme replacement which is morphemic;[3] subtrac-

[2] This is expressed as a correlation rather than as a criterion. If one were to use this statement as a criterion for inclusion in minor morpheme class, then one would count here certain particles in certain languages—Germanic languages with their prepositions, etc., for example (Hockett's 'functors'). But by formal criteria alone, particles are included with stem classes rather than with affixes and operators.

[3] In Hopi, the plural of actor is marked by affix for some verb stems and by suppletion for other verb stems. Such suppletion is an example of maximum phoneme replacement, since there is no phonemic resemblance between V_2 for *singular actor* (redundantly marked by one noun or noun substitute, N, which co-occurs with the verb) and V_2 for *plural actor* (marked by another co-occurring N). The morphemic density would be the same whether one states pïti *singular actor arrives*, as V_1, one morpheme, and ʔŏki *plural actor arrives*, as V_2, two morphemes (verb stem as one mor-

tion which is morphemic (Bloomfield's 'minus feature'); and, finally, word order which is morphemic. It is possible to encounter a fixed reduplication which is not morphemic in some languages (e.g. the verb in 'He poohpoohed it'). If not morphemic, phoneme replacement and subtraction are morphophonemic, as indicated in fn. 3. So also, when not morphemic, transformations in word order are distinguished (as allosyntactic free alternants) from transformations included under operators.

An operator functions like an affix—i.e. with reference to a particular stem class or word class—and yet an operator is itself not an affix. Recognition of the coordinate nature of affixes and operators is useful in typologizing across languages and in structuralizing within languages. In typology, it permits such statements as that goal is marked by an affix in one language, by word order in another. In structuralizing a given language, as Hopi, it permits such statements as that plural is marked by affix alone for some nouns (suffix); by operator alone for other nouns (reduplication); and, redundantly, by both affix and operator for still other nouns.

The next section, **2**, gives instructions for reading a simple but unfamiliar notation for allomorphic density. The same notation serves also for indicating morpheme class and word class. Where a single stem is included in a word, and where the particular class of the stem (as N, V, etc.), is the same as that of the word, no special notation is used. Thus, the notation N-3 indicates that a stem of noun class is followed by three suffixes, and that the sequence as a whole belongs to the noun word class. We would key English *hard* as A (adjective class stem) which remains A for word class; but in contrast we would key *examination* as V-1 – noun (i.e., *examine* is verb stem, followed by one suffix, *-ation*, with the emerging word class different than the stem class and hence specified in the notation), as in the phrase *hard examination*: A V-1 → noun. Similarly we would key *harden* as A-1 → verb; and so on. Whenever two or more different stem classes are specified in compound, the notation indicates the emerging word class.[4]

2. The allomorphic density of a given language is comparable to that of another language only when some agreement is reached as to what variations in a language are to be included, and what are to be excluded in the comparison. Decision is easy when practical considerations are involved. We are agreed that it would be extremely cumbersome and impractical to include variations due to paralinguistic influences; pharyngealization of vowels in some kinds of emotional speech is an example. Stylistic variations, as for word-in chants, are also excludable. So also are idiolectic or dialectic differences among speakers.

Some typologies go so far as to distinguish automatic morphophonemics—variations derivable from a base form by rules—from non-automatic morphophonemics. There is an interesting procedural difficulty with this distinction. If the first published version of Tübatulabal were typologized, morpheme alternants involving changes of vowel length would be counted as non-automatic; they are described by lists, with one kind of vowel change peculiar to certain sets of allomorphs and another kind peculiar to other sets of allomorphs. If the morphophonemic restatement of Tübatulabal were typologized, the alternants involving changes of vowel length

pheme and the operator, maximum phoneme replacement, as the other morpheme), on the one hand, or, on the other hand whether one states ʔǒki as a morpheme alternant of piti before *plural suffix* with the shape zero. The latter would also yield a density of two morphemes: piti + Ø = ʔǒki.

[4] Word classes may be distinguished not only by morphological criteria, as for the present examples, but also by tagmemic (or tagmetic?) criteria, to be discussed in another installment of this paper. So also, reserved for future discussion, are distinctions among affixes, as 'derivational' versus 'inflectional', and 'closing' versus 'continuing' in strings of affixes.

would be counted as automatic, since these alternants are derived from special formulae or bases (new symbols added, and morphophonemic values set up beside phonemic values for the symbols previously used)—derived from such formulae by quite complex rules. In consequence, the original structuring of Tübatulabal would yield one morphophonemic type, while the structural restatement would yield another type.

This points up a general danger to be avoided if possible—the danger of obtaining a typology of descriptions instead of a typology of particular languages. To obtain the latter, we want our typology to come out the same way, irrespective of whether it is based on some first description or some structural restatement, assuming of course that there is no contradiction between the two structuralizations (as there is certainly not in the Tübatulabal instance).[5]

So far as allomorphic density range is concerned, we obtain such desired uniformity in typology by counting in one morphophonemic category morphemes whose allomorphs are automatically changeable (as identified above), and morphemes whose allomorphs are not automatically changeable and are, accordingly, statable only by lists. Such 'changeable' morphemes are indicated by a prime after the symbols for stem or operator or affix in a particular sequence. In the case of English plural, /-s ~ -z ~ -iz/ exemplifies a set of automatic alternants; but /-in/ and other allomorphs exemplify the non-automatic alternants. Thus, merging automatic and non-automatic alternations into one general category of allomorphs, we would key *books*, *boys* and *oxen* as N-1′, but *wives* and *houses* as N′-1′. In the first set, each noun stem has a 'unique' shape (*book*, *boy*, *ox*), and is accordingly symbolized by N without a prime. In the second set, the noun stems have more than one allomorph, and hence are symbolized by N with a prime (N′). In all instances cited, the plural suffix is changeable and hence symbolized by a prime on the number denoting how many suffixes follow the N′ or N, namely N′-1′ or N-1′. When a language has two affixes after N, for example, each having more than one allomorph, this is indicated by two primes after the number giving the sum of affixes: 2″. If one of the two suffixes had more than one allomorph, and the other had a unique shape, this would be indicated by one prime after the number giving the sum of the suffixes: N-2′.

The preceding exposition is spelled out in a very elementary manner because it introduces an unfamiliar notation for marking morphophonemics. We turn now to a procedural reservation on some extended uses of this notation.

Though we cover automatic alternation by this notation, we do so only when two adjacent morphemes are involved—within or between words (internal and external sandhi)—in the span relevant to the morphophonemics. But longer spans than this are relevant to what might be called 'long span morphophonemics'—e.g., vowel harmony in Osmanli Turkish, alternation of stress in Hopi, and alternation of vowel loss in Kannaḍa. Should we use our morphophonemic notation where many of the allomorphs in a long word (as in Turkish) or in a contour bounded by pause

[5] It is easier to say what is not in contradiction between two successive structuralizations than what is contradictory. For example, where the two are isomorphic they are certainly not in contradiction. Reference to the earlier morphophonemic statement is given in Martin Joos' Readings in Linguistics, 88–92 (Swadesh and Voegelin, A Problem in Phonological Alternation, Lg. 15.1–10).

The problem discussed at this point is a general one: to make typology independent of different styles and fashions of structuralizing that produce adequate descriptions. We will attempt this for morphosyntactic density ranges by merging all classes of morphemes, or those that are distinguished in non-unique ways, in one density range; and we will report our results in subsequent installments of this paper. Though it has already been said that this is not very interesting, since it levels differentials, it nevertheless permits distinction among different descriptions of the same language which are isomorphic from those which are not.

juncture (as in Hopi and Kannaḍa) are statable in terms of a span longer than adjacent morphemes? We think not. In long span morphophonemics, the allomorphs of several successive morphemes can be stated at one stroke by indicating the onset and direction of the alternations. When taken individually, each morpheme may be found to have allomorphs, while the successive changes really constitute a single alternation over the whole span. Where such long span morphophonemics obtains, we count it as just one change for the whole span rather than as the sum of all the particular changes.

Thus a Turkish verb stem followed by a half dozen suffixes, all of them participating in a vowel harmony alternation, would be noted as (V-6)′. Hopi alternation of stress and Kannaḍa alternation of vowel loss is given by rule rather than by special notation.

3.1. Phonological subsystems have been sufficiently explored (see bibliography in fn. 1) to require no summary here. Density ranges can obviously be stated, in terms of the ordering principle followed in listing phonemes (as in our *Guide to Transcribing*, or as in Hockett's *Manual*—obstruents vs. sonorants vs. others for the latter; or the more conventional distinction in the former —stops vs. fricatives vs. nasals vs. liquids vs. semivowels). Some languages show a denser population of one kind of consonant or another. So also for vowels, the density is given according to the ordering principle of rounded-unrounded, front to back, high to low, where the inventories of particular languages are compared. So also, Series Generating Components (SGC) can be stated in terms of density ranges. Other unit phonemes, as junctures, can be stated as density ranges in respect to the length of span between pause junctures.

3.2. There is no doubt that once a number of related or unrelated languages have been structuralized, typological comparabilities can be obtained for various purposes, not only in phonological but also in other density ranges. Such typologies would be relevant to translation problems, as already indicated, and also to prediction in historical changes, to distinguishing continuity and discontinuity in the changing structure of a language family (i.e., accounting for the analogous as well as the homologous features in comparative grammar), to the language learning problem and psycholinguistics in general, to the nature of language or the 'universal grammar' problem, and finally to anthropological linguistics.

The central interest of anthropological linguistics is in the investigation of relatively unknown, unwritten, or unstructuralized languages. Its impact on linguistic theory and practice in general has been uneven. It has led us to look for the structure of each language in its own terms, with the implication of thereby discovering its uniqueness. But its discovery procedures have been utilized rather than explicitly stated. Hence it is now often said that they are unsystematic, and it is further predicted that they are incapable of being systematized.

If discovery procedures are to be systematized, the systematization would begin not with the uniqueness of structures, but in the framework of comparability that typology offers. It has already been assumed that in phonemicizing, it is most efficient to work in a context of possible phonemic distinctions derived from a knowledge of the languages of the world.[6] Typology would help to make explicit this assumption, and thereby might lead to a systematization of discovery procedures—in morphosyntax as well as in phonology.

4. Throughout these density ranges, we are not using a yes-or-no basis for typologizing. We select, as the basis for density ranges, only what is actually known as occurring with greater or less density, in all natural

[6] See p. 3 ff. and fn. 2, p. 28 in Guide to Transcribing (Anthropological Linguistics 1.6, 1959).

languages. When the minimum of occurrence is stated as zero, the statement is relevant to certain sequences in a language which also shows greater than zero in other parts of the range in question. The density ranges discussed in this paper illustrate this.

Every language known to us includes stems in some sequences, though some words in some languages occur with zero stem.

So also, every language known to us includes some affixes, but here the density range varies enormously. The occurrence of particular classes of affixes, as suffixes, may be zero in reference to all words of verb class, as in Navaho; but there still is a low density range of suffixes in Navaho words of noun class. Conversely, the occurrence of prefixes is almost zero in Eskimo, but not quite; there is one prefix in Eskimo.[7]

Instead of saying that there are languages without any prefixes, we give the density range of occurring affixes (suffixes) with zero for the prefix inventory, and hence zero for all text occurrences. In consequence, we avoid speaking of a prefix type of language or of a suffix type of language. Instead, density range typology compares languages in terms of the paucity or richness of affixes in respect to stem classes.

If, now, a language were found without any affixes, it would not be comparable to the languages actually compared in density range typology. Parallel statements are possible for operators.

This requirement—that some part of a density range have actual membership—serves as the criterion for typological utilization or non-utilization in phonology too. The usual distinction between vowels and consonants can of course be utilized, since all known languages will show actual occurrence of each.

Morphemes are phonemically distinguished by tone in some languages, but not in others. Rather than setting up a type of tone language versus a type of toneless language, density range typology would set up tone as a particular kind of series generating component (SGC). It seems safe to say that every language will include some SGC. In one language the SGC will be tones, in another stresses, in another nasalization—combinable with vowels and thereby generating unit phonemes.

[7] The prefix is tak- in sequence with stems classed as 'demonstrative-directionals' by Morris Swadesh, South Greenlandic (Eskimo), VFPA 6.36 (1946).

SIERRA POPOLUCA MORPHOLOGY

BEN ELSON

SUMMER INSTITUTE OF LINGUISTICS

0. Introduction
1. Word inflection
2. Person markers
3. Verb inflection
4. Verb derivation
5. Substantive inflection
6. Substantive derivation
7. Numeral inflection
8. Participle derivation
9. Particle derivation

0. Sierra Popoluca[1] has four INFLECTIONAL SYSTEMS and one inflectional subsystem. A SYSTEM consists of a group of affix morphemes having roughly the same privileges of occurrence. One inflectional system occurs (potentially) with all words except a few particles, the second occurs with verbs, the third with substantives, and the fourth with numerals. Four ROOT classes are distinguished: verb roots, substantive roots, particle roots, and numeral roots. Depending on their occurrence with the inflectional systems there are five STEM classes, but including internal composition as an auxiliary criterion six stem classes are distinguished: A-particles, particles, verbs, substantives, numerals, and participles.

Using other criteria it would be possible to arrive at still other classifications (for example, using participle derivation as a criterion three stem classes are obtained: verbs, nonverbs, and A-particles), but the six stem classes listed above seem to be the most useful and to correspond most closely with important syntactic clssses.

The large classes of substantives and verbs are broken down into subclasses depending on their occurrence with inflectional subsystems. In making these subclassifications the person markers play an important part.

There are three DERIVATIONAL systems: morphemes used to form verb stems, to form substantive stems, and to form participle stems.

As indicated above we distinguish between derivation and inflection, i.e., between stem formation and word formation. However, the line between these two is difficult to draw. Our procedure has been to decide more or less impressionistically which elements constitute derivation and which inflection, then to propose a definition of the difference between the two and adhere rigidly to the definition. Such definitions will be given for each stem class separately.

Inflectional morphemes, in principle, occur with all of the stems in the class which is defined by that inflectional system. However, many combinations that should be possible have not been observed; undoubtedly many simply do not occur in the language. As Wonderly[2] suggests, these "combinations do not occur because of limitations of the general grammatical structure and because of restrictions due to semantic or cultural context." We have tried to indicate where, we feel, the restrictions are clearly due to "the general grammatical structure." It is obviously impossible to indicate, in a work of this scope,

[1] This paper is a shortened version of the author's doctoral dissertation Sierra Popoluca Morphology, Cornell University, 1956 (to be published in Spanish translation by the Universidad Veracruzana, Jalapa, Ver., Mexico). Other publications on Sierra Popoluca are: Elson, Ben, Sierra Popoluca Syllable Structure, IJAL 13.13–17 (1947), Foster, George M. and Mary Foster, Sierra Popoluca Speech, Smithsonian Institution, Institute of Social Anthropology, Publication 8 (1948). Pp. 45.

[2] Wonderly, William L., Zoque I: Introduction and Bibliography, IJAL 17.3 (1951).

cases where the nonoccurrence of a given combination is due to semantic factors.

Any single-morpheme utterance consists of an A-ROOT. Different kinds of A-roots are determined by different privileges of occurrence. Morphemes which parallel A-roots in most of their privileges of occurrence but have not been observed to occur as single-morpheme utterances are B-ROOTS. There is no essential difference between A-roots and B-roots other than that given, we therefore speak simply of ROOTS.

Any morpheme that is not a root is an AFFIX. As indicated above affixes are either inflectional or derivational. A single root, or a sequence of root plus other morphemes which may occur with inflectional affixes is a STEM; derivational affixes occur with roots or stems to form other stems. A WORD is any stem which occurs in an utterance with or without inflectional affixes, or is any root which occurs in an utterance to which no affixes may be added. It would also be possible to define a word as any morpheme or morpheme sequence with which either a word suffix (see **1**), or no suffix at all, may occur.

There is a small class of particles that do not occur with any affix of any system. We call these A-PARTICLES. The following is a representative list: ídʸʌk *action not in progress*, yákiñ *almost, just now*, wʌ́ʔ·p *may, might*, má?·m *already, really*, mu *time-subordinate clause indicator*, yúli *probably*, teh *let's go!*, hére *that*.

Others are borrowings from Spanish: ni *neither, nor*, i *and*, éntre *and*, kon *with*, pórke *because*, péro *but*, kási *almost*, dónde *from*, dénde *from then, there*, kási *almost*, pára *for*, de *of, from*.

1. Except for the small class of A-particles, the affixes of the word-inflection system occur with the members of all other classes in the language. There are two classes that occur only with word-inflection affixes: particles and participles. These classes are distinguished from each other by different methods of stem composition.

The affixes of the word-inflection system are:

Order I: -nam *still*, -am/-m[3] *now, just now*, -tʸim[4] *also, just*; as in anʌ́kpatʸim *I'm going also*, yó·yatʸim *just a pig*, dʸátʸim *also no!*, heʔm óywʌ́ʔ·ptʸim *the one who also went*, pʌ́·šiʔñam *now a man*, hésʌgam *just then*, anʌ́kpam *I'm going now*, cʌ́·šiñam *still a child*, dʸánam *not yet*, nʌ́kpanam *he is still going*.

Order II: -wey *I said*, -un *he said*, -hak *Oh, I see*, -kip *you see*;[5] as in čímpawey *a dog, I said*, anʌ́kpawey *I'm going, I said*, hʌ́ʔ·wey *yes, I said*, dʸáʔun *no, he said*, á·tébedun ítʸ *it is in Soteapan, he said*, ikúʔdun sá·mñi *he ate bananas, he said*, yó·yahak *it's a pig, I see*, míñpamhak *he's coming now, I see*, dʸágip tʸí· *you're welcome*, wʌ́gip *okay, fine*, sʌ́ʔ·pkip *now, you see*, míñpakip *he's coming, you see*.[6]

2. The person marker subsystem of inflection occurs with both verbs and substantives. Its range is broader than any other inflectional system except word inflection. There are three sets of person markers. Set I: a- *first person*, ta- *inclusive person*, mi- *second person*, (third person is unmarked); Set II: an- *first person*, tan- *inclusive person*, iñ- *second person*, i- *third person*; Set III: man- *first to second person*, an$_2$- *second to first person*.

[3] -am occurs after consonants, -m occurs after vowels.

[4] It is possible that -tʸim could be analyzed as -tʸi (see **5.1.**) plus -am. However, the fact that such forms as wíñtʸitʸim *before that, first* are possible, where both -tʸi and -tʸim occur, and the fact that -tʸi does not occur with verbs (unless it is a part of -tʸim) suggest that they are two different morphemes. The meanings, however, are similar; there is undoubtedly an historical connection between them.

[5] -kip has an alternant -gip after monosyllables ending in a vowel.

[6] There is also a form -caŋ that occurs in one or two set phrases before -tʸim: wʌápcaŋtʸim ušaŋ *he can also do quite a little*.

All three sets[7] occur with both verbs and nouns (nouns are the largest subclass of substantives—other subclasses of substantives are determined mainly by the way in which they occur with the person markers).

In addition to person, every occurrence of a member of Set I marks PARTICIPANT, every occurrence of a member of Set II marks ASSOCIATE, and every occurrence of a member of Set III marks both participant and associate. If either first or second person is involved it will be marked; third person is marked only if no first or second person is present.

The participant is defined as the person most closely connected with the action or situation; the associate is one step removed from the action or situation. Set I, marking participant, appears (a) with nouns: apá·šiñ *I am a man*, tayó·mo *you and I are women*, mičá·ši *you are a child*, káwah or he? káwah *it is a horse*; (b) with intransitive verbs: anʌ́kpa *I'm going*, tahó·ypa *you and I take a walk*, mimíñpa *you come*, wí?kpa *he eats*; (c) with transitive verbs plus -tá· *indefinite performer*: ako?ctá·p *I am hit, someone hit me*, takwi?ktá·p *you and I are fed*, mi?a?-mtá·p *someone is looking at you*, ku?ttá·p *someone is eating it*; (d) with reflexive stems (these must also be inflected with -tá· *indefinite performer*): anakó?·ctá·p *I am hitting myself*, minišó·ŋtá·p *you are hitting yourself (with your fist)*, tanatá·ŋtá·p *you and I are chopping ourselves*, nakká?·tá·p *he is killing himself*.

The participant may also be marked syntactically: ko?ctá·p šíwan *someone is hitting John* or *John is being hit*, ʌč apá·šiñ *I am a man*, nʌ́kpa šíwan *John is going*, natʌ́·ŋtá· šíwan *John chopped himself*.

Where both participant and associate may be marked syntactically either the participant or the associate may be marked by the person prefixes. First or second person, in either category, takes precedence over third person; when both first and second person are involved set III is used. Such situations obtain with transitive verbs in both main and subordinate clauses:

(a) the participant is marked: akó?cpa *he hits me*, takó?cpa *he hits you and me*, mitʸóypa *he loves you*, mišóŋpa šíwan mimíč *John hits you with his fist*, kó?c *it got bruised*: (b) the associate is marked: aŋkó?·cpa *I hit him*, taŋkó?cpa *you and I hit him*, míč iŋkó?cpa šíwan *you hit John*, ikó?cpa *he hits him*.

Set II, marking associate, appears (a) with nouns: antʌ́k *my house*, taŋkáwah *your and my horse*, iŋká·ma *your cornfield*, ihá·tuŋ *his father*; (b) with intransitive verbs in subordinate clauses: mu annʌ́k a?ñiš cá·ñ *when I went I saw a snake*, dʔa wʌ?á·p iñʌ́k *he cannot go*; (c) with transitive verbs plus -tá· *indefinite performer* in subordinate clauses: míñpa iko?ctá· *someone will come and hit him*.

In subordinate clauses the use of set II with verbs where set I would otherwise be expected, appears to indicate that the meaning of 'associate' may be extended to include 'subordination of construction' in most situations.

Set III, marking both participant and associate, appears (a) with transitive verbs: maŋkó?cpa *I am hitting you*, aŋkó?cpa *you are hitting me*; (b) with nouns: manhá·tuŋ *you are my father*, a?ná·pa *I am your mother*.

3. Any SP stem which may occur with suffix -pa *incompletive* is a verb stem. Verb stems are divided into transitive or intransitive by virtue of different inflectional possibilities. Some stems participate in all the inflectional possibilities of either class and are considered members of both classes.

[7] There are obviously possibilities of further analysis (or subanalysis): a- *first person*, t- *inclusive*, mi-/i-/m- *second person*, (i- occurs with associate, m- with first person, mi- elsewhere), third person unmarked, -n-/i- *associate* (i- occurs with third person, -n- elsewhere); when associate is not present participant is indicated.

Set III can be reconstructed morphophonemically as mi- + an- > man-, and a- + iñ- > an-.

Although the above is possible, since there appears to be little gain in analytical insight, we feel the best description results from the three sets presented.

Inflectional differences between transitive and intransitive stems occur at three points in the inflectional system: (a) different person markers are used; (b) number suffixes function differently when occurring with intransitive stems than when occurring with transitive ones; and (c) different indefinite performer suffixes are used.

3.1. There are two orders of inflectional prefixes and six orders of inflectional suffixes. These affixes show certain relationships among themselves which will be mentioned as they are discussed.

Prefix order I contains a single prefix wága- *together*: wágaoyyáh *they went together*, miwáganʌktá?mpa kon pé·toh *you will go with Peter*, aŋwágako?ctá?mpa *we, together, hit him.*

Prefix order II contains the three sets of person markers (**2** above).

Suffix order I contains two number suffixes: -ta?m/-ta[8] *first and second person plural*, and -yah *third person plural*. With intransitive stems -ta?m/-ta occurs with first and second person prefixes, -yah occurs with the bare stem (or with the third person prefix in subordinate clauses) marking plural participant: anʌktá?mpa *we (excl.) go*, tawitʸtʸá?mpa *we (incl.) walk*, mipoytʸá *you all ran*, kekyáhpa *they fly*, mu ikekyáh *when they flew*. With transitive stems -ta?m/-ta occurs with first and second person prefixes to mark plural participant or plural associate depending on which is indicated by the person markers; with Set I plural participant is marked, with Set II plural associate: aŋko?ctá?mpa *we (excl.) hit him*, ako?ctá?mpa *he hits us (excl.)*. With transitive stems -yah occurs with first and second person prefixes of Set II marking plural third person participant; it indicates plural third person associate when occurring with the person prefixes of Set I. When -yah occurs with the third person prefix of Set II, it ambiguously marks plural participant or plural associate or both: aŋko?cyáhpa *I hit them*, ako?cyáhpa *they hit me*, iko?cyáhpa *he hits them, they hit him, they hit them*, mu ikó?cyahtá· *when someone hit them.*

Order II contains two suffixes: -ta· *indefinite associate* and -nʌ·m *indefinite participant*. -ta· occurs only with transitive stems; its function is to cancel the possibility of indicating the associate. When the verb is in a main clause only Set I person prefixes occur; when the verb is in a subordinate clause only Set II person prefixes occur: akó?ctamtá· *someone hit us*, ku?ttáwom *someone ate it*, mu ikó?cta aktáŋ *when someone hit him he fell*. -ta· is obligatory with reflexive stems (see **4.1**): anakó?-·ctá·p *I am hitting myself.*

The suffix -nʌ·m *indefinite participant* occurs only with intransitive stems and no person marker is present; its function is to cancel the possibility of indicating the participant. Such constructions have not been observed to occur in subordinate clauses: wá·nná·mpa[9] *someone is singing*, wí?-·kyahná·mpa *people are eating.*

Order III contains a single suffix -gak *repetitive* which may occur with both transitive and intransitive stems: akčíŋtagákum *someone bathed him again*, nʌ́kyahgákpa *they are going again.*

Order IV contains two suffixes: -to *desiderative* and -tʌ?p *frustrative* which occur with both transitive and intransitive stems. However, -to occurs only when -pa (see Order V) is present and -tʌ?p only when -pa is not present:[10] nʌktó?·ba[11] *he wants to go, is about to go*, hó·ygaktó?·ba *he wants to go for a walk again*, taŋkú?ttamtó?·ba *we (incl.) want to eat it*; a?óytʸamtʌ́?p *we*

[8] -ta occurs word final, -ta?m occurs elsewhere (but see fn. 15). The glottal catch is lost when the suffix is unstressed; this is also true of all morphemes.

[9] The vowel of the root which precedes the suffix -nʌ·m is lengthened when the root morpheme has the phonemic shape CVC (where V is any short vowel nucleus).

[10] Actually -to and -tʌ?p could be considered allomorphs: -to occurs with incompletive, -tʌ?p occurs with completive. However, we prefer to treat them as separate morphemes, admitting that the preference is probably arbitrary.

[11] The morphophonemic sequence -to + -pa yields the phonemic sequence -tó?·ba.

(*excl.*) *went in vain*, akká?·yahtá?p *they tried to kill me but failed.*

Order V contains three suffixes. One suffix -pa/-p[12] *incompletive* occurs with verb stems (both transitive and intransitive) in main clauses or in clauses introduced by the particle iga: ahó·ypa *I take a walk, I will take a walk*, miwí?kpa *you eat, you will eat*, it^{y}áŋyahgaktó?·ba *they want to chop it again*, nanʌktá·p *someone takes him.*

Two suffixes occur with verbs in subordinate clauses: -pa·m *incompletive subordinator* and -wʌ·m *completive subordinator.* When these forms occur with intransitive stems Set II person markers are used. Examples: imiñpá·m ipá?t tú·m úšpiñ *while coming along he found an alligator*, togóyum hémʌk antá?·bapá·m *he got lost there while I was fishing*, ammíñt^{y}aŋwá·m akcá?yt^{y}a túh *when we came, the rain caught us*, hésʌk iko?cwá·m kít^{y} ikʌ́ *then when he hit him he broke his hand.*

Order VI contains one suffix -um/-wom[13] *definite completive* which follows -wʌ·m of Order V:[14] sé·tum *he returned*, iñáknewá·mum *after he went.*

Two modal suffixes occur which are mutually exclusive with those listed in Orders IV, V, and VI: -iñ *obligatory* and -ʌ/-0[15] *imperative.* Constructions in which -iñ is present occasionally occur as independent syntactically, but are more frequently preceded by the particle iga *that, in order that*: anági?ñam *I should go now*, ikó?čiñ *he ought to hit him*, iñá?ma?y iga mí?ñiñ *he told him that he should come*, išunpa iga ininágiñ *he wants to take it.*

Verbs containing -ʌ/-0 *imperative* are always independent constructions. Every occurrence of -ʌ/-0 with intransitive stems indicates second person participant; with transitive stems it indicates second person associate; a first person participant may be marked in addition: pó·yʌ *run!*, miñt^{y}á?·mʌ *all of you come!*, nʌkgá·kʌ *go again!*, kó?·cʌ *hit him!*, ko?ttá?·mʌ *all of you hit him!*, ko?cyá·hʌ *hit them!*, nakó?·ctá·hʌ *hit yourself!*[16]

3.2. There are two kinds of verbal auxiliaries which should be included in a treatment of verbal inflection.

One type of verb phrase is characterized by the occurrence of a single set of inflectional affixes with the verb complex (auxiliary stem plus main stem). There is a small group of intransitive verbs that occur both as auxiliaries and as single intransitive verbs: nʌk *to go*, miñ *to come*, oy *to go and come*, moh *to begin*, yah *to finish*, ha?y *to prolong*, kus *to be satisfied.* Each of these roots may also occur with -ne (see **4.1.**). There are two verbal particles whose function parallels that of the verb roots above, except that they are uninflected: teh *come on!*, ta?·mu *let's go!*

Inflectional affixes occur as follows (the auxiliary verb always precedes the main verb): prefixes and suffixes of Orders I and II occur with the main verb, Orders IV, V, and VI, -iñ, and -ʌ occur with the auxiliary verb. When the main verb is transitive -ʌ occurs with both the auxiliary verb and the main verb. There is a suffix -i[17] *intransitive*

[12] -p occurs optionally (a) after -to (anʌktó?p or anʌktó?·ba *I want to go, am about to go*), and (b) after a suffix or dissyllabic stem ending -V(?)y (iku?dá?p or iku?dá?ypa *he eats the other's (food)*, hé·m tʌgá?p or hé·m tʌgáypa *there he is going in*).

[13] The two forms occur optionally. Some speakers prefer one form and some the other, though most speakers use both. -wom is the preferred form in slow speech.

[14] -um is undoubtedly a combination of *-u *completive* (cf. Wonderly, 1951, p. 141) and -m *now, just now* (see **1**) which has frozen as a single form.

[15] -O indicates that there is no overt affix present. It occurs optionally when a stem contains -a?y (see **4.1.**) and/or -ta?m: amácayt^{y}á?·mʌ or amácayt^{y}am *You (pl) grab it/them for me!*

The vowel immediately preceding ʌ is always lengthened if possible (i.e. unless the preceding vowel is followed by a consonant cluster): pet + -ʌ > pé·tʌ *sweep it!* but poks + -ʌ > póksʌ *double it* (i.e. *corn*) *down!*

[16] The root vowel is lengthened when the reflexive stem occurs with -ta (see **4.1.**).

[17] -i is not considered a regular part of verb inflection, discussed in **3**, because it is limited (with a few exceptions) to auxiliary constructions. We view it as being more closely related to the deriva-

main stem indicator which is obligatory with intransitive main stems (preceding Order I suffixes): nák i?a?myáh *they went to see it*, miñtʸó?·ba ma?na?mgák *I want to come and see you again*, óytʸʌ?p wí?·ki *he went there to eat in vain*, mohgákum iwát *he began to make it again*, mí·ñʌ miko·ñi *come and sit down!*, ná·kʌ pá·kʌ *go and get it!*, téh tahó·yi *come on for a walk!*

A few verbs of motion, nʌk *to go*, miñ *to come*, se·t *to return*, have been observed to occur with -i (see above) even though the auxiliary verb is not present: mí·ñi tuŋnó·m *he is/was coming in the road*, ná·ki á·tébet *he is/was going to Soteapan.*

In certain verb complexes when the main verb is a transitive stem, the derivational prefix ku- (see **4.3.**) and the imperative suffix -ʌ (see **3.1.**, **6.1.** and cf. -i discussed above) occur with the transitive stem. The auxiliary verb is inflected as described above. The meaning of this construction is similar to the meaning of the constructions where -ta· *indefinite associate* is present: nákpa ikumé?·cʌ *someone goes to look for him.* Only a few instances of this combination have been observed.

When the suffix -ta· *indefinite associate* occurs with the main stem in a verb complex the person prefixes of Set II occur (as they do when this construction is in a subordinate clause): nákpa i?a?mtá· *someone goes to look at it* or *someone goes and it is looked at*, míñpa taŋko?ctá· *someone comes to hit you and me*, but míñpa takó?c *he comes to hit you and me.*

The second auxiliary verb phrase type contains a single auxiliary stem sʌ *to be about, to walk.* It is inflected with Set II person markers only, and follows the main verb which is fully inflected. The person marker occurring with sʌ must match the one that occurs with the main verb stem: anákpa ansʌ́ *I'm going along now*, aŋkó?cpa ansʌ́ *I'm hitting him now*, wí?kpa išʌ́ *he is eating now.*

tional (see **6.1**) than to the inflectional system, but as an integral part of neither. When -i occurs the preceding vowel is lengthened if possible.

4. There are three criteria used to distinguish SP verb derivational affixes from inflectional ones: (a) any morpheme which may occur immediately following the reduplicative morpheme (**4.3**) is derivational; (b) any affix which may occur with a root in such a way that the two together (root and affix) form an immediate constituent with another root, the whole then forming a stem, is derivational; (c) any affix which may occur in more than one linear arrangement within words, relative to roots and/or other affixes, is derivational. Any affix meeting any one of these criteria is considered derivational in all of its occurrences, though in other respects it may function like an inflectional affix.

A stem which is also a root is a simple stem; a stem which consists of more than a root is a derived stem. Two basic methods of derivation are recognized: affixation and compounding. In some complex stems these two methods may overlap or be mixed together in a complicated way.

Stems are divided into three classes based on the different privileges of occurrence of the morphemes which comprise derived stems. Thus, some affixes occur only with verb roots (with a very few exceptions), others occur only with substantives and particles to form verb stems which, in some ways, then function like verb roots. We call these forms, including roots themselves, INNER FORMATIONS. Inner formations occur together to form COMPOUNDS. Compounds may also be formed by combining verb roots or stems with substantives or particles. Finally, either inner formations or compounds may occur with still another group of affixes to form new stems; these we call OUTER FORMATIONS.

4.1. Seven affixes may occur singly or in combinations with inner formations or compounds to form stems: -ne *durative*, -a?y *benefactive-indirective*, -ka *instrumental*, -o?y *indefinitizer*, ak- *causative*, na- *associative-causative*, na$_2$- *reflexive-reciprocal.*

All stems in which one of these affixes is an IC are said to be outer formations; the

affixes are outer affixes. Occasionally, these affixes function like inner affixes in that the resulting stem may occur in compounds.

The usual arrangement of outer affixes, both linear and hierarchical, is indicated in the following chart. Numbers indicate the usual order in which ICs are obtained by successive cuts.

2	6		5	4	3	1
na_2-	ak- na-	Stem	-oʔy	-ka	-aʔy	-ne

-ne/-ni[18] *durative*[19] usually functions like an inflectional suffix. It may occur with all verb stems, and is usually followed by no aspect suffix at all or by -um *definite completive* indicating that the action of the stem was accomplished in the past, but that the results continue into the present. It may, however, also be followed by -pa *incompletive*: nʌkné *he has gone*, ikuʔtné *he has eaten it*, ikoʔnnéʔ·ba *he is carrying it along on his shoulder*, ikʌ́·ŋkʌŋkó·ñkané *he has sat on it down on it fearfully*, nʌ́kpa iʔáʔ-mamné *he is going to look at it over and over*.

na_2- *reflexive-reciprocal* occurs with a large number of transitive roots and stems. When na_2- is present such stems are termed reflexive stems. When a reflexive stem occurs with number suffixes it is reflexive or reciprocal. Reflexive stems are always inflected with -ta· *indefinite associate*. The difference between na_2- and na- *associative-causative* is that when na_2- is present the accompanying root contains a long vowel. Compare the first two illustrations. natʌ́·ŋtʌ́· *he cut himself* < tʌŋ *to chop something*, natʌŋtʌ́· (ikʌ́·ma) *someone cut (his cornfield) with him* < tʌŋ *to chop something*, anakóʔ·ctʌ́· *I hit myself* < koʔc *to hit something*, minimʌ́·ctamtʌ́·p *you are grabbing each other* < mac *to grab something*, nasó·ŋyahtʌ́· *they hit each other with their fists* < soŋ *to hit something with the fist*.

-aʔy *benefactive-indirective* occurs with most transitive stems and a number of intransitive ones; the resulting stems are transitive (with a few exceptions). In general, three functions of -aʔy may be distinguished:

(a) Along with other outer affixes -aʔy occurs with some roots which are not stems in order to form stems.

(b) When -aʔy occurs with some intransitive stems, usually transitive stems are formed: iñʌʔmáʔypa *he says to* him < nʌm *to say*, ihʌʔyáʔypa he speaks to him < hʌy *to speak*, itʸógoʔyáʔy *he lost it* < togoy *to be lost, destroyed*.

(c) When -aʔy occurs with transitive stems, it indicates that the action expressed in the stem has reference to or a connection with someone or something that is not the direct participant in the action. This indirect participant, if non-third person, is marked by the person prefix, and otherwise may be understood from the cultural context, or from the non-immediate linguistic context, or may be marked syntactically in the same clause with the verb. The following statements describe the formal features relative to the occurrence of -aʔy. (1) When the person marked by the person marker (participant class) is not the direct participant -aʔy must occur. The direct participant may be any person. (2) When the syntactically marked direct participant (noun) is inflected with a possessive person marker which does not refer to the associate, -aʔy will be present in the verb stem. Examples are: iwʌ́·cʌkpa pwé·nteh, ipʌgʌ́ʔ·ypa cʌ́y *he fixes the bridge, he gets vine for it* < pʌk *to get something*, iŋkéhay wéʔ·ñʌk *he showed wasps to him* < aŋkeh *to exhibit something*, nanʌ́gaytʸa húktʌ *someone took fire to him* < nanʌk *to take something*,

[18] -ni occurs unstressed, -ne elsewhere.

[19] This suffix was considered inflectional by the Fosters. Actually, it could be handled in any of three ways: (a) as a derivational affix (it is so treated here) which has very wide distribution paralleling the distribution of inflectional affixes; (b) as an inflectional affix (as treated by the Fosters) which has wider privileges of occurrence than do other inflectional affixes, paralleling in some occurrences derivational affixes; (c) as two different morphemes, one inflectional and one derivational.

čé?ŋay šíwan ipó?kpak *he stabbed John's ribs* < ceŋ *to stab something*, nʌ́kpa mammé?ca?y immá·nʌk *I'm going there to search for your child* < me?c *to search for something*, (but) nʌ́kpa ammé?c ammá·nʌk *I'm going there to search for my child*, itʸʌ́qŋa?y ikʌ́ *he cut the other's hand* < tʌŋ *to chop something*, (but) itʸʌ́ŋ ikʌ́ *he cut his own hand*.

Occasionally, -a?y occurs twice in a stem. This indicates a second indirect participant: ananága?yá?·yʌ *take it to him for me* < nanʌk *to take something*, itʸóba?ya?y ikučí·yuh *he took his knife away from him* < top *to take something out*, ihʌ́?·ga?yá?ypa ikáwah *he leads the other's horse for him* < hʌ·k *to lead something*.

-ka *instrumental* occurs primarily with intransitive stems; the resulting stems are transitive. It generally indicates that the participant is the means, or instrument, by which the action indicated in the stem (without -ka) is performed. Often basically transitive roots occur with -o?y; then this new stem occurs with -ka: itʸʌ́?ŋoyká?·ba *he chops with it* < tʌ?ŋo?y *to be chopping*, anté·ñká?·ba *I stand on it* < te·ñ *to stand up*, imʌ́·čká?·ba *he plays with it* < mʌ·č *to play*, aŋki?ŋká *I got up on it* < ki?m *to go up*.

-o?y *indefinitizer* occurs primarily with transitive stems; the resulting stems are intransitive. It serves to take the emphasis from the 'object', implied in a transitive stem, and place it upon the action of the stem: yó·špa?dó?ypa *he will help, he is helpful* < yo·špa?t *to help someone*, cʌgó?ypa *it moves* < cʌk *to touch something*, tuhó?ypa *he will* shoot, *he does lots of shooting* < tuh *to shoot, shoot something*.

ak- *causative* occurs with a large number of stems; with intransitive stems the resulting stems are transitive; with transitive stems a second participant is added: iknʌ́kpa *he sends him* < nʌk *to go*, ikkó?cpa *he causes him to hit other* < ko?c *to hit something*, iká?·ba *he kills it* < ka *to die*, ikkuwe?-ntʌ́kpa *he causes it to flash up* < kuwe?ntʌk *to flask up*, ikwʌ́?kpa *he questions him* < wa?k *to ask for something*.

na- *associative-causative* occurs with both intransitive and transitive stems; the resulting stems are transitive; with transitive stems a second participant is added. na- generally indicates a kind of causative situation in which the associate also participates, taking a subordinate part, in the action: ininʌ́kpa *he takes it* < nʌk *to go*, ini?itʸ *he has it* < itʸ *to be, exist*, inihókspa *he causes the other to hoe it along with him* < hoks *to hoe something*, iniyó·špʌ́?tpa *he causes the other to help him with himself* < yo·špa?t *to help someone*.

Changes in the arrangement of outer affixes are possible. There are a limited number of SP verb roots that are not stems. Some of these occur with inner affixes (see **4.3**); a few occur with outer affixes to form stems. Usually stems formed in this way function as inner formations and may occur as members of compounds, or even (rarely) with inner affixes: inimótpa *he mixes it together* < mot underlying root, mʌgó?ypa *he lies* < mʌk underlying root, hʌ?dʌ́?ypa *it is carried downstream* < hʌ?t underlying root, tʌ́·né?·ba *he stands quietly* < tʌ· underlying root.

There are a few roots that are also stems but whose meanings cannot be predicted when occurring with outer affixes: aktʌ́ŋpa *it falls* < tʌŋ *to chop something*, ipakká?·ba *he knocks him down* < pak *to bump something*, ipacʌ́?ypa *he throws it away* < pac *to throw something*, mʌmné *he is ill* < mʌm *to lay something down horizontally*.

A few stems have a rearrangement of the outer affixes which occur in the stem: nanʌ́kne?ó?y *he owes* [*money*], ikó?npakne?-ʌ́?y *he carried it along on his shoulder for him*.

In the statement of the outer affixes given above successive IC cuts would first cut off -o?y then ak- at a later stage: akka?ó?ypa *he will kill* < akka *to kill something*. However, in some forms ak- is removed by the first IC cut, even though

-oʔy is present in the stem: ikmátoʔŋoy čáhka *he caused the deaf person to be able to hear* < matoʔŋoʔy *to be hearing*, ikčihóʔypa túh *he caused it to rain* < čihoʔy *to be beating*.

4.2. Any stem which contains two or more verb roots, or one or more verb roots and one or more substantive or particle stems is a compound verb stem. SP verb compounds fall into two basic types: verb-element compounds and compounds which contain a substantive or particle.

Three types of verb-element compounds are recognized: special compounds, regular compounds, and enclosed compounds.

There is a relatively small group of roots and stems larger than roots that appear to have special compounding possibilities, in that they may occur with regular compounds and enclosed compounds, as well as with roots and inner formations. Stems in which one of these special compounding forms is an IC are special compounds. Two special compounding forms occur preposed and the rest occur postposed to other stems. Many of these special compounding forms have a slightly different meaning when so occurring than they do when they occur as the nuclear element of a stem. At least one, kuru·t, is not itself a stem. Three others, poy, cak, and nas, have been observed to occur with outer formations.

The special compounding verb root poy (basic meaning: *to run, flee*) occurs preposed to a number of different stems. In this position poy means *a little while, quickly*: poyhéhpa *he rests a little while* < heh *to rest, pant*, ipóʔyaŋhó·ppakcákpa *he closes it by rolling something in front and leaves it for a little while* < aŋho·ppakcak *to close something by rolling something over it and leave it* (in this example cak is also a special compounding form, poy has a somewhat wider distribution than the other forms of this class).

The second preposed special compounding form is aŋheʔ·ga *to hurry*: aŋhéʔ·gačíkpa *I'm going to hurry and pick my corn* < čik *to pick corn*.

A few examples of the special compounding forms that occur postposed are given below. Others will be found in the full Spanish version.

nas *to do first, do just before* (basic meaning: *to pass by*): awiʔknáspa *I'll eat first* < wiʔk *to eat*, iŋwéhpaʔtnáspa *he prays to him beforehand* < aŋwehpaʔt *to pray to someone*. ket *to finish* (basic meaning: *to go down*): ihokskétpa *he finishes hoeing it* < hoks *to hoe something*. moŋ *to do in the evening* (basic meaning: *to sleep*): wiʔkmóŋpa *he is eating supper* < wiʔk *to eat*, iŋsóshakmóŋpa *he cooks a lot of it in the evening* < aŋsoshak *to cook a lot of something*. kuru·t *to turn out partly good and partly bad*: iwátkurú·t *he made it but it did not turn out well* < wat *to make, do something*, išóskurú·t *he cooked them but only part of them got done* < sos *to cook something*. aŋhak *to do for a long time* (basic meaning: *to lead, rule over someone*): tʌ́·neʔaŋhák *he stood quietly for a long time* < tʌ·ne *to be standing quietly*, moʔŋaŋhak *he overslept* < moŋ *to sleep*.

There are three classes of inner formations which form regular compounds; these are based on their occurrence in two positions within the compound stem: those occurring only in position I, those occurring only in position II, and those occurring in either position. Generally speaking reduplicated forms occur in position I. The action of the stem in position I either precedes in time the action of the stem in position II or is somehow logically prior to it.

If both stems in the compound are intransitive the resulting stem will be intransitive; if one of the stems in the compound is transitive the resulting stem will usually be transitive. However, compounds containing aʔm *to look at something* and an intransitive root are intransitive: aʔŋkíʔmpa *he looks up* < kiʔm *to go up*, aʔmnʌ́kpa *he looks into the distance* < nʌk *to go*.

Regular compounds of root plus root: itʸʌŋmácpa *it pecks something* < tʌŋ *to chop something* and < mac *to grab some-*

thing, ihʌypáʔtpa *he addresses him* < hʌy *to speak* and < paʔt *to find something*, iʔišpʌ́kpa *he is acquainted with him* < iš *to see something* and < pʌk *to take something*, ihóʔ·yíšpa *he hates him* < hoʔy *to be angry* and < iš *to see something*.

Regular compounds of inner formations (larger than root) plus root: iŋwehpáʔtpa *he prays to him* < aŋweh *to shout* and < paʔt *to find something*, ičʌ́gaŋhéʔk *he frightened him* < cʌk *to touch something* and < aŋheʔk *to be startled*, itʸúʔcapákpa *he follows him* < tuʔca· (cf. tuʔc *tail*) and < pak *to move horizontally against something, to bump something*.

Regular compounds of inner formation plus inner formation or outer formation: wó·neʔaktúŋ *he lay down* < wo·ne *to be lying down* and < aktʌŋ *to fall*.

A few compounds contain compounds as one of the constituent parts: itʸʌ·mwʌ́·cákpa *he stretched it into place* < tʌ·m *to stretch something* and < wʌ·cak *to fix something* (from wʌ· *good* and cak *to leave something*).

There are roots which occur in position II and form an IC with a prefix which is preposed to the root in position I. The two together (prefix and root in position II) then constitute one IC and the root in position I the other. These prefix-root combinations are ENCLOSING FORMS, and the stems that are formed are ENCLOSED COMPOUNDS. The prefixes used are the inner prefixes aŋ- and ku- (see **4.3**) and the outer prefix ak- (see **4.1.**). A few examples are presented below; others will be found in the Spanish version.

ku-..-aʔy *to do completely* (kuʔaʔy) *to become clean, clear*): kunʌ́gayyáh *they all went away* < nʌk *to go*, kučiksáʔypa *he itches all over* < čiks *to itch*. aŋ-..-aʔy *to open a space by ...* (aʔŋaʔy *to open something*): iŋpiʔŋáʔypa *he clears a space by picking up some* < piŋ *to pick up something*, iŋcecáʔy *he chips off a small piece from the tree* < cec *to plane, shave wood*. aŋ-..-nuʔk *to close by ...* (aŋnuʔk *to close, close something*): iŋtaʔnnúʔkpa *he closes it by making a fence* < taʔn *to fence*. aŋ-..-hak *to do a great deal of* (aŋhak *to lead, rule something*): iŋsoshákpa *he cooks a great deal so there will be enough* < sos *to cook something*, iŋhʌ́·shákpa *he imagines it* < hʌ·s *to think of something*. ku-..-ket (kuket *to go down to someone*): ikuyoʔykét *he jumped down on it* < yoʔy *to jump*, kucʌ́goykétpa *there is shaking down everywhere* < cʌgoʔy *to move*.[20]

Nonverb-verb compounds are of several types. A very productive pattern is that in which the syntactic participant of a transitive verb is incorporated into the verb stem itself; the new stem thus formed is intransitive. Such a syntactic participant must, in every case, be a noun: ká·mtʌŋpa *he cuts a cornfield* < tʌŋ *to chop something* and < ká·ma *cornfield*, mokméʔcpa *he searches for corn* < meʔc *to search for something* and < mok *corn*, sʌkñíppa *he plants beans* < ñip *to plant, plant something* and < sʌk *beans*, káwahaknʌ́ʔkpa *he waters the horse* < aknʌʔk *to cause to drink* and < káwah *horse*, tʌ́gakté·ñpa *he builds a house* < akte·ñ *to cause to stand* and < tʌk *house*.

Other nonverb-verb compound patterns are much less productive. They include (a) noun-verb combinations: ahʌ́·šitʸogóy *I got confused* < togoy *to be lost* and < hʌ́·ši *thoughts*, kó·sté·ñpa *he kneels* < te·ñ *to stand up* and < kó·so *knee*, ikó·bakpíhpa *he annoys him* < pih *to be hot* and < kó·bak *head*, imá·sanwátpa *he blesses him* < wat *to do, make something* and < má·san *Spaniard, white man*; (b) 'adjective-like' nouns with verbs: iwʌ́·cákpa *he fixes it* < cak *to leave something* and < wʌ· *good*, iyʌkkʌ́ŋpa *he paints it black* < kʌŋ *to paint something* and < yʌk *black*; (c) two special compounding substantives: téget *responsibility, fault*, and kʌ *hand:* tégetnʌ́kpa *he must go* < nʌk *to go*, tégetwítʸpa *he limps* < witʸ *to walk*, ikʌ́ʔ·piŋpa *he tracks it* < piŋ *to pick up something*, kʌ́ʔ·pút *he escaped* < put *to go out*, ikʌ́ʔ·húypa *he hires workers* < huy *to buy something*; (d) certain particles:

[20] This is the only bimorphemic enclosed stem (cʌk *to touch something*, -oʔy *indefinitizer*) encountered.

ičá·mhó$^{?}$·yíšpa *he hates him very much* < ho$^{?}$·yiš *to hate someone* and < ca·m *very*, amédʸohá·šitʸogóy *I got partly confused* < hʌ·šitʸogoy *to get confused* and < medʸo *half*.

4.3. In addition to the roots themselves there are three types of inner formations: stems with inner affixes, verbs derived from nonverbs, and uniquely derived stems.

There are two inner prefixes, ku- and aŋ-, which differ from outer affixes in that they rarely form an IC with more than a single root. These prefixes, ku- and aŋ-, are mutually exclusive, and aŋ- is also mutually exclusive with ak- and occurs rarely with na-.

The core of meaning of the prefix ku- seems to be *otherness*. This basic notion manifests itself in two ways: a distributive-repetitive connotation suggesting other occurrences of the same action in time or space, and a 'separation' from the normal routine of action the root would suggest with a 'specialization' which may involve another person. With either of these meanings ku- is still mildly productive.

ku- occurs (a) with some roots that are not stems: kunu·m *to stalk something*, kumušuŋ *to swim under water*, kucat *to send someone*; (b) with stems which denote the action has broad spatial dimensions; most of these are verbs derived from nouns by the affix -a·: kupi$^{?}$·ča· *to get dark* < pi$^{?}$·čʌ *darkness*, kupʌ$^{?}$ma· *to be covered with weeds* < pʌm *weeds*, kuhe·he$^{?}$a· *to be surrounded with mosquitos* < he·he *mosquitos*, kukeh *to dawn* < keh *to appear, be visible*, kutogoy *to be covered (as by tall grass)* < togoy *to be lost*, kumoŋ *to sink* < moŋ *to sleep*; (c) to denote that the action is repeated, often in different places: ku$^{?}$a$^{?}$m *to look for something* < a$^{?}$m *to look at something*, kute·ñ *to stop repeatedly* < te·ñ *to stand up*, kuho·p *to roll about* < ho·p *to roll*, kupiŋ *to sort something* < piŋ *to pick up something*, kutʌŋ *to test by chopping something* < tʌŋ *to chop something*; (d) to denote a specialization of action related to another person: kuhʌy *to speak evil of someone* < hʌy *to speak*, kutʌgʌy *to defend someone* < tʌgʌy *to enter*, kunu$^{?}$k *to go up to someone* < nu$^{?}$k *to arrive*, kuhoks *to hoe at another's place* < hoks *to hoe something*, kupʌk *to believe, obey someone* < pʌk *to take something*.

The core of meaning of the prefix aŋ- seems to be *togetherness*. Derived from this basic idea are such notions as 'concentration of action to a point', 'motion toward', 'closing', that are present in most of the stems in which aŋ- occurs. In general, aŋ- appears to be somewhat less productive than ku-.

aŋ- occurs (a) with some roots that are not stems: aŋsah *to open one's mouth*, aŋmat *to speak, visit*, a$^{?}$ŋa· *to fight, argue*, aŋya$^{?}$k *to be sad*; (b) to denote the action of 'closing': aŋmiñ *to heal* < miñ *to come*, aŋnu$^{?}$k *to close, close something* < nu$^{?}$k *to arrive*, aŋpah *to close a door or gate* < pah *to pen up something*; (c) to indicate motion of objects toward each other or of one object toward another in a fixed position: a$^{?}$ŋa$^{?}$m *to wait for someone* < a$^{?}$m *to look at something*, aŋte·ñ *to lie in wait for someone* < te·ñ *to stand up*, aŋnʌk *to chip metal* < nʌk *to go*; (d) to indicate concentration of action to a point: aŋhʌy *to sound, bang* < hʌy *to speak*, aŋtʌŋ *to chip off a piece of bark* < tʌŋ *to chop something*, aŋmeh *to sharpen something* < meh *to rub something*; (e) with an intensive, often purposive, meaning: aŋweh *to shout* < weh *to weep*, aŋkeh *to exhibit something* < keh *to appear, be visible*, aŋhak *to lead, rule someone* < hak *to cross, cross something*, aŋtʌgʌy *to work hard* < tʌgʌy *to enter*.

Inner suffixes include a directional suffix and the reduplicative suffix.

-ʌ$^{?}$yʌ$^{?}$y/-ʌ$^{?}$y[21] *direction along a line* occurs with a few roots. When this suffix is present no other derivational affixes have been observed to occur: yá$^{?}$m púdʌ$^{?}$yá$^{?}$y *here he went out* < put *to go out*, yám

[21] -ʌ$^{?}$yʌ$^{?}$y occurs with monosyllabic roots, -ʌ$^{?}$y with dissyllabic roots.

cúkuʔmáʔy *here he started out this way* < cukum *to begin, start.*

Almost every SP root and a few larger forms may occur with the reduplicative suffix which consists of a repetition of the phonemic shape except that, in single syllable roots, · (vowel length), ʔ and stress do not occur in the reduplicated form. The reduplicative morpheme means *repeated action.*

The sequence of root plus reduplicative morpheme is never immediately inflected; it must have with it another stem or an outer affix (-oʔy or -ne): íšišóʔypa *he is able to see* < iš *to see something*, nʌ́mnʌmnéʔ·ba *he keeps saying it over and over* < nʌm *to say*, yʌ́ksayʌ́ksanéʔ·ba *he goes like this* < yʌksa· *to do like this* (< yʌks *like this*), itʸʌ́ŋtʌŋhákpa *he cuts it all apart by chopping* < tʌŋ *to chop something* and < hak *to cross, cross something*, ikuhʌ́·thʌdáʔypa *he marks it all up* < hʌ·t *to mark something* and < ku-..-aʔy *to do completely.*

Certain roots occur only with the reduplicative morpheme: múrumúruneʔ·ba *he is visiting*, tʸékšektʸékšekneʔ·ba *he wiggles his lower jaw*, póytʸʌkpóytʸʌgóʔypa *he runs around in a large circle* (cf. póy *to run*).

Two suffixes -a· *factive* and -ʌʔy *possessive* occur with nonverb stems to form verb stems. However, three nonverb stems occur with outer affixes: ceʔmóʔy *he urinated* < cem *urine*, tʸiʔñóʔy *he defecated* < tʸiñ *feces*, súksukkáʔ·ba *he is shivering* < suksuk *cold.*

-a· *factive* occurs with a number of nonverbs to form verb stems. The resulting stem is always intransitive: mʌhá·p *he gets big* < mʌh *big*, tʌgá·p *he builds a house* < tʌk *house*, mogá·p *it produces corn* < mok *corn*, hóʔ·gá·p *it smokes* < hoʔko *smoke*, wʌdʸaʔyá·p *he gets old* < wʌdʸá·ya *old man.* There are a number of forms that appear to have verbal meaning and do not occur as substantives, but which must occur with -a·: yóʔ·šá·p *he is working*, kóʔ·wá·p *he is injured*, kupóʔ·dʸá·p *he gets a haircut.*

-ʌʔy *possessive* occurs with a large number of nonverbs (mostly nouns) and the resulting stems are intransitive, though some of them may also be transitive. The intransitive stems are usually not inflected by other than the person markers and number suffixes. The intransitive stems indicate that the person marked possesses the item indicated by the nonverb stem: tʌgáʔy *he has a house* < tʌk *house*, kuʔyáʔy *he has wood* and ikuʔyʌʔypa *he puts wood on it* < kuy *wood*, tuʔŋáʔy *it has a road* and itʸuʔŋáʔypa *he follows him* < tuŋ *road*, káʔ·máʔy *he has a cornfield* < ká·ma *cornfield*, ahémʌgáʔypa *I'm going over there* < hémʌk *over there.*

There are a number of unique derivations, part of which can be identified as a morpheme, but the other part of which does not occur elsewhere. These forms usually are two syllables and different vowels occur in each syllable (see also two-syllable roots below). A partial list of these forms follows: mo·was *to yawn*, cf. was *to bite something*; čiput *to turn inside out*, cf. put *to go out*; honʌm *to do something in a certain way* [tanhonʌ́mpa *how shall we do it* ʔ], cf. nʌm *to say*; sapitʸ *to hem something*, cf. pitʸ *to make a cage-like framework with string.* Some two-syllable roots may be similarly related to single syllable roots: aŋmatamatane *to be visiting*, cf. aŋmat *to speak*; wokowokone *to rattle*, cf. wok *to bark*; hʌkʌhʌkʌʔoʔy *to follow along*, cf. hʌ·k *to pull something*; kuwʌhʌʔaʔy *to dissolve*, cf. wʌh *to disappear.* These are not merely reduplicated alternants since each of the short forms may be reduplicated as well: wokwokne *to bark repeatedly.*

All observed verb roots contain either one or two syllables. Single syllable roots have the following shapes: CV·: pʌ· *to get fat*, pa· *to have diarrhea*; CV: pa *to strain something*, no *to burn something*; CVC: pet *to sweep, sweep something*, nʌk *to go*; CV·C: se·t *to return*, hʌ·k *to permit something*, kaʔm *to stay placed*; CVʔ·C: hiʔ·š *to rub*

something with the finger, me$^{?}$·s *to wall up something*; CVCC (final C is always /s/): poks *to double down corn*, heps *to dish out something*; CV$^{?}$CC: so$^{?}$ps *to get tired*, ne$^{?}$ks *to tie grass on a house roof*. In the above patterns the first C means a consonant or nothing: CVC: uk *to drink*.

Two-syllable roots all contain the same vowel in both syllables: CVCV: šit^{y}i *to shiver*, ʌlʌ *to move back and forth*; CVCVC: cukum *to start*, togoy *to be lost*, tʌgʌy *to enter*, hukun *to stir something*; CVCCVC: sehken *to become loose*, t^{y}kšek *to move the lower jaw*; other patterns: wo$^{?}$ŋko *to roll about on the ground*, potoks *to bend a finger joint*, taŋga *to be upside down*.

5. Substantives are considered to be a single class because they exhibit certain distributional characteristics in common: (1) they all participate to some extent in the substantive-inflection system, (2) verbs are derived from all members of the substantive class in the same way, (3) participles are derived from all members of the substantive class in the same way. Substantives are divided into major subclasses chiefly by their occurrence with the person markers.

5.1. The inflectional system occurring with nouns consists of the person markers (see **2**) and four orders of inflectional suffixes. Nouns are differentiated from other classes of substantives in that they (nouns) occur optionally with Set II person markers. Other classes either do not occur with Set II person markers, or a member of Set II is obligatory.

There follows a chart of noun inflection. All classes of substantives participate in this system in one way or another.

Per. markers	Order I	Order II	Order III	Order IV[22]
Set I				
a-		-ma		
ta-				
mi-	-aŋhoh	-ho·m		
Set II				
		-aŋho·m	-tam	
an-	Stem	-kʌ$^{?}$·m		
tan-				
iñ-		-yukmʌ	-yah	
i-		-kukʌ$^{?}$m·		
Set III				
	-tam	-aŋku$^{?}$k		
man-			-gak	-t^{y}i
an$_{2}$-				

[22] Cf. fn. 4.

The meanings of these forms will be given as each group is discussed. The occurrence of person markers with nouns has already been discused in **2.**

Some arbitrary limitations of distribution have been observed: -aŋhoh has been observed to be followed only by -ho·m, but by no other suffix; when -tam occurs in Order I only -aŋhoh and -kʌ$^{?}$·m have been observed following it; -gak occurs either immediately after the noun stem or preceded by -kʌ$^{?}$·m, -gak is usually followed by -t^{y}i, in noun inflection -t^{y}i has not been observed without -gak.

The various orders of suffixes are discussed in a more or less random sequence since discussion in terms of general categories is more helpful than one in terms of orders.

SP nouns are inflected with two number suffixes, -tam and -yah, which have counterparts in verb inflection, and with the suffix -aŋhoh.

-yah *plural* (Order III) occurs with most nouns; however, the lack of this suffix does not necessarily mean singular. It is possible to speak of more than one without using any number suffix. Plural affixes are optional. tʌ́kyah *houses*, kúyyah *trees*, kó·ñkuyyah *chairs*, yó·yayah *pigs*, cʌ́·šiyah *children*.

-aŋhoh *mass plural* (Order I) occurs with a number of nouns: tʌgáŋhoh *many houses together, a village*, tó$^{?}$·dáŋhoh *much paper in a pile*, caáŋhoh *many rocks, a rocky place*.

-tam *plural* (Order I) occurs with a few nouns referring to human beings: yó·mtam *women*, cʌ́·št^{y}am *children*, it^{y}ʌ́·wʌtam *his brothers*, wó·ñt^{y}am *little girls*. This subclass may also be pluralized by -yah: wó·ñiyah *little girls*. A different subclass, containing a

limited number of stems, whose inflection is otherwise identical with that of nouns, is pluralized with -tam. This subclass contains 'adjective-like' words: wʌ́tam *good ones*, cábactam *red ones*, yúkmʌtam *high ones*.

-yah and -tam are also used to pluralize the person markers (see also **3.1**). When Set II is present -tam occurs with first and second person markers and -yah occurs with the third person marker. When used in this way -tam occurs in Order III:[23] antáktam *our (excl.) house*, aŋká·mhó·mtam *in our (excl.) cornfield*, taŋkáwahtam *our (incl.) horse*, iyo·yayah *their pig, his pigs, their pigs*, -yah occurs with Set II person markers but pluralizes the noun rather than the person: antákyah *my houses*.

-tam also occurs with Sets I and III when these occur with a noun stem: tapá·šiñtʸam *we all are men*, miyó·yatam *you are all pigs*, mancá·štʸam *you are my children*. -yah occurs optionally in some of these situations: (heʔ·) pʌ́·šiñyah *they are men*, miyó·yayah *you are pigs*.

There are seven locative suffices in SP (Order II). Nouns inflected with these affixes do not function syntactically as subject or object, nor do they occur with Sets I and III person markers.

-mʌ *instrumental*: kúymʌ *with a stick, wood*, ká·pmʌ *with an arrow*, petkúymʌ *with a broom*, nʌ́k pó·yí·mʌ *he went running*. With a few nouns -mʌ has locative function: itʸʌ́kmʌ *his country* cf. tʌk *house*, ítʸ iʔé·ʔ·smʌ *he is in bed* cf. céʔ·s *bed*.

-ho·m *locative, in*: tʌkhó·m *in the house*, ká·mhó·m *in the cornfield*, tó·thó·m *in the book*. -ho·m also occurs with some nouns which, when occurring with verbs of motion (*go, come*, etc.), could be glossed *to get . . .*: nʌ́kne mokhó·m *he has gone to get corn* cf. mok *corn*, óyyah sʌkhó·m *they have been to get beans* cf. sʌk *beans*.

-aŋhó·m *locative, among*: yó·yaʔaŋhó·m *among the pigs*, itʸʌ́·wʌtaʔmaŋhó·m *among his brothers*.

-kʌʔ·m *locative, at*: nʌ́ʔ·kʌ́ʔ·m *at the water*, ikʌ́·mkʌ́ʔ·myah *at their cornfield*, tʌkkʌ́ʔ·m *at the house*.

-yukmʌ *locative, up, upon*: mé·sahyúkmʌ *upon the table*, tʌkyúkmʌ *on the house*.

-kukʌʔ·m *locative, under*: kó·ñkuykukʌ́ʔ·m *under the chair*, mé·sahkukʌ́ʔ·m *under the table*.

-aŋkuʔk *locative, in the midst*: pʌ́·šiʔñaŋkúʔk *in the midst of the people*.

Suffixes -gak *also, again* and -tʸi *just* very often occur together with noun stems. Words containing both are most frequent in the following kind of dialogue: tʸí·sʌ́ʔ·ba iŋwʌ́t *what are you doing?* tʌ́kgaktʸi sʌ́ʔ·ba aŋwʌ́t *I'm building a house*. -gak may occur without -tʸi: anúʔkta nʌ́ʔ·kʌ́ʔ·ŋgak *we arrived at the next water*.

5.2. Pronouns are differentiated from nouns in that they occur with Set I person markers but not with Sets II and III. Pronouns are also inflected for number, occur with a few locative suffixes, and with suffixes -gak and -tʸi. Five pronouns are distinguished: ʌč *I, me*, mič *you*, he/heʔm[24] *he, him, she, her, it*, peʔm *that one*, iʔ *who, whom*. Set I person markers occur with pronouns as intensives (when the marker for the same person is added), and to form stative constructions (when the marker for a different person is added): aʔʌ́č *I*, taʔʌ́č *you and I*, mimíč *you*; ahé *I am he*, miʔí· *who are you?*

Number suffixes occur with pronouns: ʌ́čtʸam *we, us*, míčtʸam *you (pl.)*, íʔtʸam *who (pl.)*, héʔ·yah *they*, péʔmyah *those*.

Three locatives have been observed with pronouns: ʌčkʌ́ʔ·m *at my place*, mimíčtʸaʔ-

[23] It should be pointed out that -tam and -yah are observed following members of Order II only when functioning in this way. It would, therefore, be possible to include -yah as a member of Order I and say that Order I pluralizes the noun and Order III pluralizes the person markers, except that no member of Order II has yet been observed following -yah.

[24] Pronouns may modify nouns in syntactic constructions. Therefore, heʔm (usually translated *the*) is considered to be an alternant of he in attributive position.

maŋhó·m *among you*, peʔmyúkmʌ *upon that one*.

-gak and -tʸi occur with pronouns: ʌčgaktʸi *I also*, héʔ·gaktʸi *he also*, ĺ·gagam *who else?*

5.3. There is a group of words which are distinguished from other substantives in that they always occur with one member of Set II person markers; we call these forms nominals. Nominals do not occur systematically with any other substantive inflection. Some nominals occur with locatives, number suffixes, and -gak and -tʸi while others do not. Of the various subclasses of substantives, nominals exhibit the least unity of syntactic function: itʸú·mpʌy *all of them*, tantú·mpʌy *all of us*, ikuyúkmʌ *the reason for it*, ikusáŋwiñ *above it*, ikukáʔ·m *beneath it*, ikunášwiñ *down below it*, anhó·doŋ *my knowing*, ihóʔ·ya *his anger*.

5.4. Adjectives, like pronouns, occur only with Set I person markers and with number suffixes, but unlike pronouns, they do not occur with locatives: aháʔyaŋtam *there are many of us*, šútʸuyah *they are small*, asiŋkoh *we are five*.

6. As with verbs the difference between derivational and inflectional elements of substantives is not completely clear. The following criteria are used to distinguish them: (a) Any morpheme which occurs with a verb stem to form a substantive is derivational; (b) any morpheme which occurs as a prefix to the second stem or as a suffix to the first stem of a compound is derivational in all occurrences; (c) any affix morpheme which radically changes the meaning of the stem is derivational.

6.1. Derivational affixes occur with roots and stems of various classes to form substantive stems.

There follow illustrations of affixes occurring with substantive roots and stems to form new stems:

-tu·ku *old, valueless* occurs with a number of nouns: tʌktú·ku *old, dilapidated house*, puktú·ku *cloth* cf. pú·ki *cotton*, yuktú·ku *orphan* cf. haʔyuk *younger sibling*, pétkuytú·ku *old broom*.

-tʸa·ka *without, -less* occurs with a few nouns, generally body parts: káʔ·tʸá·ka *armless*, puytʸá·ka *without a foot*, tuʔctʸa·ka *without a tail*.

-te·roˢ *doer, performer* (see below for occurrence also with verb stems): yó·mté·ro *ladies' man*, yó·yaté·ro *hog dealer*, káwahté·ro *good horseman*.

-ko·m *place of* occurs with a few substantives: wʌ́·kó·m *good place*, tʌ́·škó·m *Ocosotepec*, čiŋkó·m *place of pine trees*.

-pak *arroyo*[26] (cf. pakus *arroyo*): ká·ŋpak *Tilapa River* (lit. *jaguar arroyo*), é·špak *Sochapan River* (lit. *crawfish arroyo*), nʌ́ʔaŋnó·ypak an unnamed (on maps) tributary to the Guasuntlan River (lit. *water-flower arroyo*).

There follow illustrations of affixes occurring with verb roots and stems to form substantive stems:

-kuy *instrumental* occurs with a large number of verb stems to form nouns: pétkuy *broom* < pet *to sweep, sweep something*, háykuy *pencil, pen* < hay *to write*, íškuy *eye* < iš *to see something*, séʔtsedóʔ·ykuy *frying pan* < seʔtsedoʔy *to be frying*. -kuy also occurs, in fewer forms than above, with meaning *intended object of action*: yóʔ·mʌ́ʔykuy *one who should marry* < yoʔ·mʌʔy *to have a wife, take a wife*, kóʔckuy *one who should be hit* < koʔc *to hit something*.

-i *result of action* occurs with a large number of verb stems to form nouns: pí·hi *heat* < pih *to heat, get hot*, háʔ·pi *corn meal dough* < haʔp *to grind something*, ú·ki *a drunk* < uk *to drink, drink something*, ʌ́kši *shelled corn* < ʌks *to shell corn*, hú·ki *cigar, cigarette* < huk *to smoke*.

-ʌ/-a *result of action* occurs with a few stems instead of -i or, in a few cases, in addition to -i: púʔ·kʌ *tropical ulcer* <

[25] This suffix has an atypical phonemic shape. It may be there has been Spanish influence.

[26] See also Wonderly, William L. Zoque Place Names, IJAL 12.217-228 (1946), esp. p. 223.

puʔk *to rot*, mó·ya *flower* < moy *to foam, flower*, kuyšáʔ·mʌ *splinter* < saʔm *to splinter off*.

-kʌʔ·wiñ *plural actors* occurs with a few stems, usually intransitive ones: kuyú·hkáʔ·wiñ *learners, pupils* < kuyuh *to learn*, yó·šakáʔ·wiñ *workers* yo·ša· *to work*.

-aŋ occurs with one verb stem to form a noun: hípsaŋ *live coal* < hips *to burn*.

-te·ro *performer of action* (see above for occurrence with substantive stems): hʌytʸé·ro *talker, one who talks a lot* < hʌy *to talk*, yó·šaté·ro *worker, one who works hard* < yo·ša· *to work*, mʌ́·čtʸé·ro *player, one who likes to play* < mʌ·č *to play*, ecté·ro *dancer* < ec *to dance*.

-pʌy *measure* occurs with a few verbs to form substantives té·ñpʌy *the height of a man* < te·ñ *to stand up*, aŋcénpʌy *the amount in a sack to where the mouth can be tied* < aŋcen *to tie something closed*. -pʌy also occurs with substantives: itʸú·mpʌy *all of it* < tu·m *one*, yʌkspʌy *like this, about this much* < yʌks *like this*.

-kʌy occurs with a few forms to derive adjectives: híkskʌy *quickly, rapid* < hiks *to hurry*, táhkʌy *fast*, šéhkʌy *light weight*, wéskʌy *fragile*.

There are several derivational prefixes which occur with substantives (and numerals) to form substantive stems:

ku- *spread, extent* (see also **4.3**) occurs with some roots that do not otherwise occur: kutó·tʸi *brains*, kuʔʌ́·ši *foolish*, kukú·ma *ring*, ikusʌ́·kʌ *his nakedness*, kupóʔk *runt*. ku- also occurs with other substantive stems and roots: kutʸiñ *lazy* < tʸiñ *excreta*, ikuwíñ *his form* < wíñ *width*, ikuyúkmʌ *the reason for it* < yúkmʌ *up, height*, ikusʌ́ŋwiñ *up above it* < sʌ́ŋwiñ *upward*. ku- also occurs with numerals and adjectives form nominals: ikutú·ku *the three of them*, ikuwʌ́stʌk *both of them*, ikutʸu·m *by himself*, ikusiŋkoh *the five of them*.

aŋ- *point* (see also **4.3**) occurs with some roots which do not otherwise occur: aŋkéhke *chin*, aŋwí·ši *beard*. aŋ- also occurs with other substantives: aŋtʌ́c *tooth of a comb* < tʌc *tooth*, aŋkí·ñi *point of a knife, stick, etc.* < kí·ñi *nose*, aʔŋáy *page of a book* < ay *leaf*, aʔŋá·ñi *pimple* < á·ñi *tortilla*.

na- *similar* (see also **4.1**) occurs with a few forms: inimí·šitʸʌ́·wʌ *his fellows Oaxaqueño* < mí·ši *Oaxaqueño* and < tʌ́·wʌ *brother, fellow*, iniyó·mtʌ́·wʌ *her sister with her* < yó·mtʌ́·mʌ *sister*.

ak- (see **4.1**) occurs with one noun: akcá *testicles* < ca *rock*.

waga- (see **3.1**) occurs with a few forms: wágayó·mo *twin girls* < yó·mo *woman*, wágapʌ́·šiñ *twin boys* < pʌ́·šiñ *man*.

6.2. Two types of compound substantives are recognized, regular compounds and enclosed compounds.

Examples of regular compounds are: nuʔnkúypak *needles* < núʔnkuy *sewing instrument* and < pak *bone*, iškúypʌk *eyebrow, eyelash* < íškuy *eye* and < pʌk *body hair, feathers*, sóhtʌm *acorns* < soh *oak tree* and < tʌm *fruit*, mʌ́ʔ·tá·tuŋ *father-in-law* < mʌ́ʔ·t *son-in-law* and < há·tuŋ *father*, kúyham *ashes* < kuy *wood* and < ham *lime*, kobáktʌk *hat* < kó·bak *head* and < tʌk *house*, cuʔé·ší *tarantula* < cu *night* and < é·ši *crawfish*, kuykó·bak *tree stump* < kuy *tree, wood* and < kó·bak *head*. There are three special compounding forms: nunta- *real, very*, naps- *pure*, and mu ma- *whole, complete*. Examples are: núntawʌ́· *very good*, nuntánʌ *real water, just water*, núntahʌ́·yi *Popoluca language*; nápsnʌ *pure water*, napshéckuy *just combs*; mú·mahá·ma *all day*, mú·macú *all night*.

There are remnants of enclosed compounds: kuči ʔŋáŋnʌ *water that has been bathed in* < čiŋ *to bathe* (cf. nʌ *water*), kuceʔáŋnʌ *water that has been washed in* < ce *to wash something*. There is an enclosing form meaning *season of*: kuhá·msʌŋ *dry season* < há·ma *sun* (cf. sʌŋ *sky*), kusʌgáŋkʌ *bean season* < sʌk *beans* (cf. áŋkʌ *yard*), kukúytʸʌʔmáŋkʌ *avocado season* < kúytʸʌm *avocado*, kumogáŋkʌ *corn season* < mok *corn*, kutúhkʌ *rainy season* < tuh *rain* (here only kʌ is present, this is perhaps an alternant of áŋkʌ.).

There is one enclosing form that appears to consist of a prefix-suffix combination: ikuhágaŋ *a piece of it* < hak *to break crosswise, go across,* ikucáʔyaŋ *something left over of it* < cʌʔy *to remain.*

6.3. There are a few forms which recur but to which no meaning can be assigned: -či: tʌ́kči *cockroach* (cf. tʌk *house*). kí·ñči *nasal catarrh* (cf. kí·ñi *nose*), pukúnči *chigger,* néʔñči *rooster's comb,* juʔŋí·či *rainbow*; -tʸi/-ñi[27]: póʔčtʸi *curly,* káʔčtʸi *blind,* tučtʸi *humming bird,* sá·mñi *banana,* sʌ́ŋñi *daylight* (cf. sʌŋ *sky*); -tuk: aŋwíñtʸuk *on the other side* (cf. wiñ *width*), pí·ñtʸuk *deserted* (cf. pi·ñ *to fade into the distance*).

There are a few unique forms in what appear to be compounds: cúʔmʌk *last night* (cf. cu *night*), hʌ́·pak *corncob* (cf. pak *bone*), náʔ·piñ *blood* (cf. nʌ *water, liquid*; also cf. Proto-zoquean *pʌn *man*), cʌ́·nhon *wild turkey* (cf. hon *bird*).

6.4. There are a few substantives derived from verbs by shift of stress: kápun *barrow* (cf. ikapúnpa *he castrates, emasculates it*),[28] iwátnas *his image* (cf. iwatnáspa *he makes it first*), aŋhákket *slope* (no verb counterpart has been observed, but the constituent parts can be identified as verbs).

6.5. SP contains a large number of substantives formed by reduplication of a root: púnpun *slick,* cáncan *the smell of oil or metal,* máymay *happy,* kíʔñkiñ *hairy, fuzzy,* tí·ttit *mestizo,* sáʔ·sa *stem*; háyaháya *fast, quick,* nútunútu *bumpy, rough skin,* čí·ščí·ši *stink bug*; íšpʌgíšpʌk *friend,* tʌ́ʔŋguytʸʌ́ʔŋguyčáncan *the smell of ink.*

6.6. Substantive roots occur in a number of different phonemic shapes: CVC: tuh *rain,* mok *corn,* hon *bird*; CVʔC: muʔk *grass,* cuʔc *corpse*; CV·C: po·k *corn stalk,* ca·ñ *snake*; CVʔ·C: koʔ·ŋ *basket,* sʌʔ·p *now*; CVʔV: máʔa *deer,* náʔa *chewing gum*; CVCV: kútʸu *harelip,* tʸáka *chick,* šútʸu *small.* A very large number of substantive roots have the shape CV·CV. Normally, the second vowel is the same as the first: yó·mo *woman,* hé·he *mosquito,* hʌ́·hʌ *house fly,* tó·to *paper,* ká·ma *cornfield,* kí·ñi *nose,* kó·ya *rabbit,* yó·ya *pig.* Other shapes are possible: óhyah *sister-in-law,* áʔ·wi *fishhook,* kʌ́ʔak *huarache,* á·tébet *Soteapan.*

7. Numerals are a very small class consisting only of the numbers two through five, the question word *how many,* and one or two other forms, yet they have a characteristic inflection which is different from all other inflectional systems in SP. Numerals occur with Set I person markers, with number suffixes (substantive inflection -tam and -yah), and with -gak and -tʸi. Numerals are distinguished from other classes by the inflectional affixes -te·n and -na·: wʌsté·n *two,* túkuté·n *three,* máktasté·n *four,* mosté·n *five,*[29] hú·té·n *how many?* The difference between -te·n and -na· is not clear. The former is usually used in counting and when the numeral is subject or object; the latter usually occurs when the numeral is in attributive position, though both forms have been observed in all positions. Forms containing -na· are frequently accompanied by a number suffix: wʌsná·yah *there are two,* atúkuná·tam *there are three of us.* Other illustrations: mihú·ná·tam *how many are there of you?,* wʌsté·ngaktʸi óy *two of them went.*

8. Members of any class of stems which participate in word inflection (see **1**) may be used for participle derivation. Two methods of participle derivation are possible: one method is used to form participles from verbs, the other to form participles from nonverbs. Since a difference in syntactic function also parallels the resultant types of participles, we distinguish V-participles and N-participles.

[27] -ñi occurs after nasals, -tʸi occurs elsewhere.

[28] The stem is of Spanish origin.

[29] Numbers above five in Sierra Popoluca have not been observed, though there are reportedly old people who know them. (I could not seem to find any who could or would give them to me.) Actually, numbers for four and five are seldom used in speech, Spanish numbers being preferred instead.

8.1. There are two suffixes used to derive V-participles; these parallel the contrast between 'completive' and 'incompletive' which is found elsewhere in the SP inflection systems: -paʔ·p *incompletive participlizer* and -wʌʔ·p *completive participlizer*. Syntactically V-participles function as participant or associate or as a part of certain noun phrases; they are usually preceded by heʔm *the, that*: heʔm nʌkpáʔ·p ipáʔtpa *the one who goes will find it*, añyó·špáʔtum heʔm aktʌŋwáʔ·p *I helped the one who fell down.*

Inflectional prefixes and Orders I and II of inflectional suffixes may occur with the verb stem before the participlizing suffixes are added: heʔm aʔñíštʸaŋwáʔ·p póy *the one we saw ran*, heʔm kóʔctawáʔ·p *the one someone hit.*

Actually the participlizing affixes appear to form an IC with the entire clause: heʔm nʌkwáʔ·p á·tébet matʌ́k togóy tuŋhó·m *the one who went to Soteapan yesterday got lost on the road*, heʔm ikoʔcwáʔ·p itʸʌ́·wʌ číʔ·tʸa kastígo *the one who hit his brother was punished.*

8.2. A single suffix, -pʌk, is used to derive N-participles: cábacpʌk *red ones* < cábac *red*, yó·mpʌk *female ones* < yó·mo *woman*, heʔm wʌsté·npʌk *the two* < wʌsté·n *two*.

Inflectional prefixes and suffixes may occur with substantive stems before -pʌk is added: ká·mhó·mpʌk *one from the cornfield*, itʸʌkkáʔ·mpʌk *the one from his house*, cábactampʌk *red ones*, kócʌkyúkmʌpʌk *one from the mountain.*

When the negative particle dʸa occurs with a verb and a participle is formed, -pʌk occurs with dʸa rather than -paʔ·p or -wʌʔ·p with the verb stem: heʔm dʸápʌk nʌ́k *the one who did not go.*

9. Certain stems which do not participate in any of the inflection systems other than word inflection, and therefore can be regarded as particles, consist of more than one morpheme.

Several stems contain an element -ʌk: hémʌk *over there* (cf. hé·m *there*), hésʌk *then*, wíñʌk *long ago.*

There are two affixes which occur with numerals and some other forms to derive particles, -mʌ *days hence* and ma- *days past*: maktásmʌ *four days hence*, tú·kmʌ *three days hence*, wʌstákmʌ *day after tomorrow*, hóymʌ *tomorrow*, amtʸʌ́ymʌ *next year*; matʌ́k *yesterday*, mawáški *two days ago*, matú·ki *three days ago*, maʔamtʸʌ́ymʌ *last year.*

Some forms in which the suffix -kʌy (see **6.1**) occurs are particles: wáskʌy *twice* <wʌs- *two*, wá·tʸkʌy *many times* < wá·tʸi *many.*

Other forms have constituent morphemes that are present in substantive derivation and inflection: aŋwíñtʸuk *on the other side*, aŋsʌ́·kmʌ *outside*, aŋhákket *slope*, aŋkʌ́ʔ·m outside (i.e. *not in the house*).

SENECA MORPHOLOGY III: EXPANDED PRONOMINAL PREFIXES*

WALLACE L. CHAFE

BUREAU OF AMERICAN ETHNOLOGY, SMITHSONIAN INSTITUTION

9. The nonsingular subjective prefix
9.3. The pronominal constituent
9.7. The number morpheme
9.10. With the modal prefix
9.12. Summary
10. The objective prefix
10.5. The singular objective prefix
10.9. With the modal prefix
10.11. The nonsingular objective prefix
10.14. With the modal prefix
10.16. Summary
11. The transitive prefix
11.2. Combinations of first and second persons
11.4. Masculine subject and object
11.7. Restricted third person subject and object
11.11. Other combinations
11.14. With the modal prefix
11.16. Summary

9.1. The term PRONOMINAL PREFIX will be used from this point on to include the singular subjective prefix described in **6**, which is the irreducible pronominal prefix, and the expansions of it to be described in **9–11**. These include the nonsingular subjective prefix (**9**), the singular and nonsingular objective prefix (**10**), and the transitive prefix (**11**).

9.2. The singular subjective prefix described in **6** may be expanded by the suffixation to it of a NUMBER MORPHEME to form a NONSINGULAR SUBJECTIVE PREFIX. The number morphemes introduced in this section are the DUAL (-ni-/-n-/-ny-) *two* and the PLURAL (-wa-/-wæ-/-wɛ-/-w-/-y-/-0-/-ti-/-n-) *three or more*. A third, the NONSINGULAR (*two or more*), will appear in **10** and **11**. These morphemes are not in all cases simply added to the pronominal morphemes of **6**, but condition certain modifications of the pronominal constituent. The following diagram shows in a general way the structure of a verb or noun containing the nonsingular subjective prefix; the broken lines indicate that the modal prefix may or may not be present (**8**):[1]

	Verb or Noun		
Modal Prefix	Nonsingular Subjective Prefix		Verb or Noun Stem
	Pronominal Constituent	Number Morpheme	

An example is ha|ti- *masculine|plural* in hatí:hsa:s *they're looking for it*, hatí:waji:-yæ? *their family*.

9.3. Instead of being added directly to the first person morpheme the number morpheme occurs with the EXCLUSIVE PERSON construction, consisting of the first person morpheme (-k-) preceded by the exclusive person morpheme (?a-), which indicates that reference is made to someone in addition to the speaker, but excluding the hearer: ?akhníyɛtha? *we (excl. du.) are hitting it* (with h between the k and n by **3.9**), ?akwáyɛtha? *we (excl. pl.) are hitting it*.

9.4. There is also an INCLUSIVE PERSON morpheme (h-/t-), indicating that the hearer is included, which occurs alone (not in construction with the first person morpheme) before the number morpheme. The inclusive allomorph is h- before the dual morpheme, t- before the plural: hniyétha? *we (incl. du.) are hitting it*, twayétha? *we (incl. pl.) are hitting it*.

9.5. The second person (s-/j-) and masculine

[1] As noted in **8**, the modal prefix only occurs in a verb, and then only with a punctual or imperative verb stem.

* Seneca Morphology I and II, IJAL 26.11–22, 123–129

ne (h-/ha-/hɛ-) morphemes occur directly before the number morpheme. The second person allomorph is s- everywhere except before the plural allomorph -0-, where it is j- (**9.8**): sniyétha? *you (du.) are hitting it*, swayétha? *you (pl.) are hitting it*. The masculine allomorph is h- before the dual morpheme, ha- before the plural allomorph -ti-, and hɛ- before the plural allomorph -n- (**9.9**): hniyétha? *they (masc. du.) are hitting it*,[2] hatíyɛtha? *they (masc. pl.) are hitting it*, hɛnís-tyɔ? *they (masc. pl.) dwell*.

9.6. Neither the feminine nor the neuter morpheme occurs before the number morpheme, but their meanings are combined in the NONMASCULINE morpheme (k-/wa-/wɛ-), which does so occur. The nonmasculine allomorph is k- before the dual morpheme, wa- before the plural allomorph -ti-, and wɛ- before the plural allomorph -n-: khniyétha? *they (nonmasc. du.) are hitting it*, watyjɛtha? *they (nonmasc. pl.) are hitting it*, wɛnís?tyɔ? *they (nonmasc. pl.) dwell*.

9.7. The dual allomorphs occur as follows: -ni- occurs before a morpheme whose basic allomorph begins with a consonant: hniyɛ:thoh *let's (incl. du.) plant!* (-yɛthw-), sni:awi? *you (du.) are carrying something* (-'[h]-awi-); -n- occurs before a morpheme whose basic allomorph begins with a vowel other than a or ɛ: ?ihne? *we (incl. du.) are walking* (-e-), khnɔhe? *they (nonmasc. du.) are alive* (-ɔhe-); -ny- occurs before a morpheme whose basic allomorph begins with a or ɛ: ?akya:tɔh *we (excl. du.) say it* (-atɔ-), hnyɛtó:s *they (masc.) du.) are waving it* (-ɛtó-).

9.8. The plural allomorphs can be divided into those which occur after the first person, inclusive person, and second person morphemes (-wa-/-wæ-/-wɛ-/-w-/-y-/-o-), and those which occur after the masculine and nonmasculine morphemes (-ti-/-n-). The first group occurs as follows: -wa- occurs before a morpheme whose basic allomorph begins with a consonant, but not before (C)æ or (y)ɔ: ?akwáthe?tha? *we (excl. pl.) are pounding* (-thé?t-), twa:nékeæh *let's (incl. pl.) take a drink!* (-'[h/:]neké-); -wæ- occurs before a morpheme whose basic allomorph begins with (C)æ: swǽ:?thɛh *(you [pl.]) climb!* (-[C]ǽ?thɛ-); -wɛ- occurs before a morpheme whose basic allomorph begins with (y)ɔ or i, and in the latter case the i is lacking: ?akwɛɔkwɛh *we (excl. pl.) have taken it for our own* (-[y]ɔkw-), swɛ?tyɔ? *you (pl.) dwell* (-í?tyɔ-); -w- occurs before a morpheme whose basic allomorph begins with a, e, or ɛ, and before the morpheme -i- *say*: ?itwe? *we (incl. pl.) are walking*, ?iswih *(you [pl.]) say it!* (-i-); -y- occurs after the first and inclusive person morphemes before a morpheme whose basic allomorph begins with o or ɔ: ?akyɔ:he? *we (excl. pl.) are alive*, tyɔhe? *we (incl. pl.) are alive*; -o- occurs after the second person morpheme (j-) before a morpheme whose basic allomorph begins with o or ɔ: jɔhe? *you (pl.) are alive*.

9.9. After the masculine and nonmasculine morphemes the plural allomorphs occur as follows: -ti- occurs before a morpheme whose basic allomorph begins with a consonant: hatíthe?tha? *they (masc. pl.) are pounding*, watíthe?tha? *they (nonmasc. pl.) are pounding*; -n- occurs before a morpheme whose basic allomorph begins with a vowel: hɛnɔ:-tɔh *they (masc. pl.) say it* (-atɔ-; **3.14**), wɛnɔ:-tɔh *they (nonmasc. pl.) say it*.

9.10. When the indicative morpheme precedes a nonsingular subjective prefix, its word-initial allomorph is wa- before the exclusive person and masculine morphemes, ?o?- before the nonmasculine, and ?e- before the inclusive and second persons. The optative allomorph is ?ae- before the inclusive and second persons, ?aa- otherwise. The future allomorph is in all cases ?ɛ-.

9.11. After any modal prefix the masculine allomorph -':- occurs in place of the word-initial h- (before the dual morpheme): wa:ni:yɛt *they (masc. du.) hit it*, ?á:ni:yɛt *they (masc. du.) might hit it*, ?ɛ:ni:yɛt *they (masc. du.) will hit it*. The initial masculine allomorphs ha- and hɛ- (before the plural morpheme) occur without the initial h and

[2] In word-initial position this prefix is homonymous with the inclusive dual (**9.4**).

the immediately preceding vowel is basically strong (*cf.* **8.7**): wa:ti:yɛt *they (masc. pl.) hit it* (with the masculine allomorph : by **3.19**), waɛnɔté?ko? *they (masc. pl.) fled* (-até?kw-), ?á:ti:yɛt *they (masc. pl.) might hit it*, ?aɛnɔ́te?ko? *they (masc. pl.) ought to flee*, ?ɛɔti:yɛt *they (masc. pl.) will hit it* (**3.14**), ?ɛ:nɔte?ko? *they (masc. pl.) will flee*. After the optative and future morphemes, but not the indicative, the exclusive person allomorph is -ya-, rather than the word-initial ?a- (**9.3**): ?a:yakwa:yɛt *we (excl. pl.) might hit it*, ?ɛyakwayɛt *we (excl. pl.) will hit it*. With these exceptions, the same allomorphs occur after the modal prefix morphemes as in word-initial position.

9.12. The following is a list of the nonsingular subjective prefixes, exemplified both in word-initial position and after the indicative and future morphemes. The verb root is -kɛ- *see* in the iterative (-kɛ́h) and punctual (-kɛ́?) aspects.

?akhni-/?akhn-/?akhny- (/-yakhni- etc.) *we (excl. du.)*: ?akhni:kɛh *we see it*, wa?ákhni:kɛ? *we saw it*, ?ɛyákhni:kɛ? *we'll see it.*

?akwa-/?akwæ-/?akwɛ-/?akw-/?aky- (/-yakwa- etc.) *we (excl. pl.)*: ?akwa:kɛh *we see it*, wa?akwakɛ? *we saw it*, ?ɛyakwakɛ? *we'll see it.*

hni-/hn-/hny- *we (incl. du.)*: hni:kɛh *we see it*, ?ehni:kɛ? *we saw it*, ?ɛhni:kɛ? *we'll see it.*

twa-/twæ-/twɛ-/tw-/ty- *we (incl. pl.)*: twa:kɛh *we see it*, ?etwa:kɛ? *we saw it*, ?ɛtwa:kɛ? *we'll see it.*

sni-/sn-/sny- *you (du.)*: sni:kɛh *you see it*, ?esni:kɛ? *you saw it*, ?ɛsni:kɛ? *you'll see it.*

swa-/swæ-/swɛ-/sw-/j- *you (pl.)*: swa:kɛh *you see it*, ?eswa:kɛ? *you saw it*, ?eswa:kɛ? *you'll see it.*

hni-/hn-/hny- (/-'ni- etc.) *they (masc. du.)*: hni:kɛh *they see it*, wa:ni:kɛ? *they saw it*, ?ɛ:ni:kɛ? *they'll see it.*

hati-/hɛn- (/-'ati- etc.) *they (masc. pl.)*: hati:kɛh *they see it*, wa:ti:kɛ? *they saw it*, ?ɛɔti:kɛ? *they'll see it.*

khni-/khn-/khny- *they (nonmasc. du.)*: khni:kɛh *they see it*, ?o?khni:kɛ? *they saw it*, ?ɛkhni:kɛ? *they'll see it.*

wati-/wɛn- *they (nonmasc. pl.)*: wati:kɛh *they see it*, ?o?wati:kɛ? *they saw it*, ?ɛwɔti:kɛ? *they'll see it.*

10.1. Both the singular (**6**) and nonsingular (**9**) subjective prefixes may be modified by the OBJECTIVE morpheme (?a-/-a-/-æ-/-ɛ-/-0-/-o-/-aw-/-ɔ-/-wa-/?ɔ-/-yɔ-/) to form a singular or nonsingular OBJECTIVE PREFIX.

Modal Prefix	Verb or Noun		
	Objective Prefix		Verb or Noun Stem
	Subjective Prefix	Objective Morpheme	

The position of the two constituents varies according to the morphemes involved; the objective morpheme sometimes precedes, sometimes follows, and sometimes occurs within the other constituent, whose membership also varies in some instances (in the nonsingular objective prefix) from that described in **9**.

10.2. The grammatical meaning of this construction varies according to the type of stem with which it occurs. With a descriptive verb stem, the pronominal constituent is modified by the objective morpheme to indicate either the performer or goal of whatever is denoted by the stem, and a neuter goal or, as the case may be, performer is understood: sanyo:h (-nyo- *kill*), with the second person singular objective prefix sa-, means either *you've killed it* or *it's killed you* (cf. **5.6**); ?akí?tyɔ?, with the first person singular objective prefix ?ak-, means either *I've seated it* or *it's seated me* (cf. ki?tyɔ? *I dwell, etc.* in **6.2**). With the masculine and feminine morphemes, however, a performer that is masculine or feminine, respectively, may also be understood: honyo:h *he's killed it, it's killed him*, but also *he's killed him*; konyo:h *she's killed it, it's killed her*, but also *she's killed her*. There are certain verb stems with which only an objective prefix occurs, and with

[3] See fn. 1.

these in the descriptive aspect the pronominal constituent is usually translated as a person or thing described: hotá?ɔh *he's asleep* (-itá?-), honya:kɔh *he's married* (-nyak-), hoyɔ:ti:h *he's smiling* (-yɔ[:]ti-).

10.3. With a verb stem in any other aspect, the pronominal constituent of the objective prefix denotes simply the goal of whatever is denoted by the stem, with either a neuter, or, under the conditions noted above, a masculine or feminine performer understood: sanyos *it's killing you*, honyos *it's killing him, he's killing him*, konyos *it's killing her, she's killing her*. With stems that take only the objective prefix, however, the pronominal constituent is usually translated as the performer: hotá?was *he sleeps* (-itá?w-; **5.3**), honya:s *he marries*, sayɔ:tih (*you*) *smile!*.

10.4. With a noun stem, the pronominal constituent is modified by the objective morpheme to indicate in most cases the alienable possessor of whatever is denoted by the stem (cf. **7.2**): ?akáɔhta? *my interpreter* (cf. káɔhta? *my ear*), sa?áshæ? *your basket* (-?ásh[æ]-), hoɛnɔ? *his song* (-ɛn-), koya:a? *her bag* (-ya[æ]-; **5.4**), ?o?nɔ? *its arrow* (-'?n-). A few stems denoting body parts occur with an objective prefix: ?akyɔ:skwa? *my rib* (-yɔskw-), ?akéthwɛhsa? *my liver* (-thwéhs-). The neuter objective prefix also occurs with many stems to indicate specific existence: ?osnɔ? *the bark* (cf. kasnɔ? *bark*), and a number of stems occur only with the neuter objective form: ?onéɔ? *corn* (-né-).

10.5. The pronominal morphemes of the singular objective prefix are identical with those of the singular subjective prefix; i.e. the first person (-k-/-ke-), second person (s-), masculine (h-), feminine (k-), and neuter (?-) morphemes. The pronominal constituent comes first unless it is the first person morpheme: s|a- *second person|objective* in sáihsa:s *it's looking for you*, h|o- *masculine|objective* in hóihsa:s *it's looking for him*, but ?a|k- *objective|first person* in ?akíhsa:s *it's looking for me*.

10.6. The first person allomorphs are -k-/-ke- after the objective, conditioned as described in **6.4**. With the first person morpheme the objective allomorph is ?a- in word-initial position: ?aka:wɛh *it belongs to me* (-awɛ-), ?akéthe?tɔh *I've pounded it*.

10.7. After the second person morpheme (s-) the objective allomorph is -a- before a morpheme whose basic allomorph begins with a consonant, o, or ɔ, but not before (C)æ or (y)ɔ: sayéhtɔh *you've hit it*, sáosæɔh *you've sprinkled it* (-osé-); -æ- before a morpheme whose basic allomorph begins with (C)æ: sé:?thɛ:h *you've climbed it*; -ɛ- before a morpheme whose basic allomorph begins with (y)ɔ or i, and in the latter case the i is lacking: sɛɔkwɛh *you've taken it for your own*, sɛ:tah *go to sleep!*; and -0- before a morpheme whose basic allomorph begins with a, e, or ɛ: se?ɔh *you have intended it* (-e?-), sa:wɛh *it belongs to you.*[4]

10.8. After the masculine (h-), feminine (k-), and neuter (?-) morphemes the objective allomorph is -o- before a morpheme whose basic allomorph begins with a consonant, a, or i, and in the latter two cases the a or i is lacking: hoyéhtɔh *he's hit it*, ko:wɛh *it belongs to her*, ?otá?ɔh *it's asleep*; -a- before a morpheme whose basic allomorph begins with or or ɔ: háosæɔh *he's sprinkled it*; and -aw- before a morpheme whose basic allomorph begins with e or ɛ: hawe:?ɔh *he has intended it*. There are, however, some roots in ɛ before which -o- occurs, the ɛ then being absent. Such roots will be listed with initial (ɛ): -(ɛ)nóhtɔ-*know* (?akénɔhtɔ? *I know it*, honóhtɔ? *he knows it*.

10.9. When a modal prefix is present, the word-initial indicative allomorph is ?- before the first person, ?e- before the second person, and wa- before the masculine, feminine, and neuter singular objective prefixes. The optative allomorph is ?ae- before the second person singular objective prefix, ?aa- before the others. The future allomorph is ?ɛ- before all of them.

10.10. After the indicative prefix (?-) the

[4] See Seneca Morphology II, fn. 11.

objective allomorph before the first person morpheme is -ɔ-: ʔɔkyɛt *it hit me*; after the optative and future it is -wa-: ʔa:wakyɛt *it might hit me*, ʔɛwɔ:kyɛt *it will hit me* (**3.14**). The second person allomorph remains s- in the singular objective prefix after any modal prefix (cf. **8.7**): ʔesa:yɛt *it hit you*, ʔaesayɛt *it might hit you*, ʔɛsa:yɛt *it will hit you*. After any modal prefix the h of the masculine allomorph is lacking and the preceding vowel is basically strong: wa:oyɛt *it hit him*, ʔá:oyɛt *it might hit him*, ʔɛ:oyɛt *it will hit him* (**3.18**). After the indicative morpheme (wa-) the feminine allomorph is -ʔak- and the neuter -ʔ-: waʔako:yɛt *it hit her*, waʔo:yɛt *it hit it*. After the optative (ʔaa-) and future (ʔɛ-) morphemes the feminine allomorph is -yak- and the neuter -y-: ʔɛyako:yɛt *it will hit her*, ʔɛyo:yɛt *it will hit it*.

10.11. As noted in **10.1**, the membership of the pronominal constituent in the nonsingular objective prefix is not identical with that of the nonsingular subjective prefix. For one thing, the objective prefix makes no distinction between exclusive and inclusive persons, but combines their meanings by prefixing the objective morpheme (ʔɔ-) directly to the construction *first person|number morpheme* (dual or plural): ʔɔkhníyɛtha? *it's hitting us* (*du.*), ʔɔkwáyɛtha? *it's hitting us* (pl.).

10.12. The objective morpheme occurs in a zero allomorph in the second person nonsingular objective prefix: sniyétha? *it's hitting you* (*du.*), swayétha? *it's hitting you* (*pl.*).[5]

10.13. In the masculine and nonmasculine nonsingular objective prefixes no distinction is made between dual and plural; i.e. the dual and plural morphemes do not occur, but their meanings are combined in the NONSINGULAR morpheme (-ti-/-n-) *two or more*: hotíyɛtha? *it's hitting them* (*masc.*), ʔotíyɛtha? *it's hitting them* (*nonmasc.*). Here the objective morpheme (-ɔ-) occurs between the masculine (h-) or nonmasculine (ʔ-) morpheme and the nonsingular. The latter occurs in the allomorph -ti- before a morpheme whose basic allomorph begins with a consonant, and -n- before a morpheme whose basic allomorph begins with a vowel (cf. **9.9**): hotíthe?tɔh *they* (*masc.*) *have pounded it*, ʔotíthe?tɔh *they* (*nonmasc.*) *have pounded it*, honóte?kwɛh *they* (*masc.*) *have fled*, ʔonóte?kwɛh *they* (*nonmasc.*) *have fled*.

10.14. When the indicative morpheme precedes a nonsingular objective prefix, its word-initial allomorph is wa- before the objective, masculine, and nonmasculine morphemes and ʔe- before the second person. The optative allomorph is ʔae- before the second person, ʔaa- elsewhere. The future allomorph is ʔɛ-.

10.15. After any modal prefix the masculine allomorph h- is lacking and the immediately preceding vowel is basically strong: waoti:yɛt *it hit them* (*masc.*), ʔáoti:yɛt *it might hit them* (*masc.*), ʔeoti:yɛt *it will hit them* (*masc.*) (**3.16**). After the optative and future morphemes, but not the indicative, the objective allomorph is -yɔ- rather than the word-initial ʔɔ-, and the nonmasculine allomorph is -y- rather than ʔ-: ʔa:yɔkwa:yɛt *it might hit us* (*pl.*), ʔɛyɔkwayɛt *it will hit us* (*pl.*), ʔa:yoti:yɛt *it might hit them* (*nonmasc.*), ʔɛyoti:yɛt *it will hit them* (*nonmasc.*). With these exceptions, the same allomorphs occur after the modal prefix morphemes as in word-initial position.

10.16. The following is a list of the singular and nonsingular objective prefixes:

ʔak-/ʔake- (/-ɔk- etc. and -wak- etc.) *me*: ʔake:kɛh *it sees me*, ʔɔke:kɛʔ *it saw me*, ʔɛwɔke:kɛʔ *it will see me*.

sa-/sæ-/sɛ-/s- *you*: sa:kɛh *it sees you*, ʔesa:kɛʔ *it saw you*, ʔɛsa:kɛʔ *it will see you*.

ho-/ha-/haw- (/-'o- etc.) *him*: ho:kɛh *it sees him*, wa:okɛʔ *it saw him*, ʔɛ:okɛʔ *it will see him*.

ko-/ka-/kaw- (/-ʔako- etc. and -yako- etc.) *her*: ko:kɛh *it sees her*, waʔako:kɛʔ *it saw her*, ʔɛyako:kɛʔ *it will see her*.

ʔo-/ʔa-/ʔaw- (/-yo- etc.) *it*: ʔo:kɛh *it sees it*, waʔo:kɛʔ *it saw it*, ʔɛyo:kɛʔ *it will see it*.

ʔɔkhni-/ʔɔkhn-/ʔɔkhny- (/-yɔkhni- etc.)

[5] These forms are thus homonymous with the second person nonsingular subjective prefix (**9.5**).

us (*du.*): ʔɔkhni:kɛh *it sees us*, waʔɔ́khni:kɛʔ *it saw us*, ʔɛyɔ́khni:kɛʔ *it will see us.*

ʔɔkwa- /ʔɔkwæ- /ʔɔkwɛ- /ʔɔkw- /ʔɔky- (/-yɔkwa- etc.) *us* (*pl.*): ʔɔkwa:kɛh *it sees us*, waʔɔkwakɛʔ *it saw us*, ʔɛyɔkwakɛʔ *it will see us.*

sni-/sn-/sny- *you* (*du.*): sni:kɛh *it sees you*, ʔɛsni:kɛʔ *it saw you*, ʔɛsni:kɛʔ *it will see you.*

swa-/swæ-/swɛ-/sw-/j- *you* (*pl.*): swa:kɛh *it sees you*, ʔɛswa:kɛʔ *it saw you*, ʔɛswɔ:kɛʔ *it will see you* (**3.14**).

hoti-/hon- (/-ʼoti- etc.) *them* (*masc.*): hoti:kɛh *it sees them*, waoti:kɛʔ *it saw them*, ʔɛoti:kɛʔ *it will see them.*

ʔoti-/ʔon- (/-yoti- etc.) *them* (*nonmasc.*): ʔoti:kɛh *it sees them*, waʔoti:kɛʔ *it saw them*, ʔɛyoti:kɛʔ *it will see them.*

11.1. The pronominal prefixes so far described often have transitive meanings (or at least translations), usually with either the performer or the goal neuter (**6.2**, **10.2**), but sometimes with two masculines or two feminines (**10.2**). There is, however, also another expanded pronominal prefix, the TRANSITIVE PREFIX, in which each of the constituents contains a pronominal morpheme; one constituent, the SUBJECT, denotes the performer and the other, the OBJECT, denotes the goal of whatever is indicated by the verb stem. A transitive prefix almost never occurs with a noun stem.[6] The following diagram indicates the general structure of verbs containing this construction:

	Verb		
	Transitive Prefix		Verb
Modal Prefix	Subject	Object	Stem

An example is s|k- *second person|first person* in skihsa:s *you're looking for me.* The order of the constituents varies, depending on the particular morphemes involves, and in some cases the object contains the objective morpheme, thus defining another construction subclass (**2.15**) of the objective prefix described in **10**. In addition, certain of the transitive prefixes occur as constituents with the number morphemes, defining another subclass of the transitive construction.

11.2. First to be considered are the transitive prefixes containing a first person subject and second person object or vice versa. The constituents are simply the first person (k-/-ke-) and second person (s-/ta-/-ɔ[:]-/-ɔ[:]y-) morphemes, and the subject precedes the object: kɔyétha? *I'm hitting you*, skyɛtha? *you're hitting me.* As subjects, the first and second person allomorphs are k- and s- respectively, except that in a word that ends in the imperative morpheme the second person allomorph is ta- before the first person: takɔh *give it to me!* (-ɔ-). As an object, the first person morpheme has the allomorphs -k-/-ke-, conditioned as described in **6.4**. The second person morpheme has, as an object, the allomorphs -ɔ(:)- and -ɔ(:)y- (**4.6**), with -ɔ(:)- occurring before a morpheme whose basic allomorph begins with a consonant or i (and the i is lacking), and -ɔ(:)y- occurring before a morpheme whose basic allomorph begins with a vowel other than i: kɔ?tyɔ? *I've seated you*, kɔya:wi:h *I've given it to you* (-awi-).

11.3. The two constitutes described above also occur followed by the dual and plural morphemes, in which case the second person allomorph is -0- as the object, and the allomorphs of the number morphemes are identical with those described for their occurrence after the first person and inclusive person morphemes in **9.7–8**. The grammatical meaning here is unique: the number morpheme sets a limit on the number of individuals denoted by either of the constituents. With the dual morpheme either the subject or the object, or both, refers to a group of two, but neither can refer to a larger group: khniyétha? *I'm hitting you* (*du.*), *we* (*du.*) *are hitting you* (*sg.*), or *we* (*du.*) *are hitting you* (*du.*); skhniyétha? *you* (*sg.*) *are hitting us* (*du.*), *you* (*du.*) *are hitting me*, or *you* (*du.*)

[6] An example of such an exceptional occurrence is shetwa- *we* (*incl. pl.*) . . . *him* in shetwáhsɛ:nɔ? *he is famous or notorious among us*, with the noun stem -hsɛnɔ́? *name, fame.*

are hitting us (*du.*). With the plural morpheme at least one of the constituents refers to a group of three or more: kwayétha? *I'm hitting you* (*pl.*), *we* (*du.*) *are hitting you* (*pl.*), *we* (*pl.*) *are hitting you* (*sg.*), *we* (*pl.*) *are hitting you* (*du.*), or *we* (*pl.*) *are hitting you* (*pl.*); skwayétha? *you* (*sg.*) *are hitting us* (*pl.*), *you* (*du.*) *are hitting us* (*pl.*), *you* (*pl.*) *are hitting me*, *you* (*pl.*) *are hitting us* (*du.*), or *you* (*pl.*) *are hitting us* (*pl.*).

11.4. The masculine morpheme (h-/-0-/sh-/she-/he-) occurs as the first constituent in the transitive prefix before any of the subjective or objective morphemes or constructions (**6**, **8**, **9**) that contain the first, inclusive, or second person morphemes, as well as before the UNRESTRICTED THIRD PERSON morpheme (-ɔwɔ-/-ɔwæ-/-ɔwɛ-/-ɔwɔy-/-ɔw-)[7] *he*, *she*, *they*, the feminine singular objective construction, and the NONSINGULAR THIRD PERSON objective construction. If the second constituent is objective (contains the objective morpheme), it is the object; if subjective, it is the subject and the masculine morpheme is the object. But the masculine morpheme, whether subject or object, always comes first.

11.5. Before the first person singular object the masculine allomorph is h- and the objective allomorph -a- (cf. **10.6**): hakyétha? *he's hitting me*. Before the second person singular object the masculine morpheme has a zero allomorph in word-initial position and the second person allomorph is -y- (cf. **10.7**): yayétha? *he's hitting you*. Before the first person dual and plural objects the masculine allomorph is sh- and the objective allomorph -ɔ- (cf. **10.11**): shɔkhníyɛtha? *he's hitting us* (*du.*), shɔkwáyɛtha? *he's hitting us* (*pl.*). Before the second person dual and plural objects (**10.12**) the masculine allomorph is she-: shesníyɛtha? *he's hitting you* (*du.*), sheswáyɛtha? *he's hitting you* (*pl.*). The maculine allomorph sh- also occurs before the feminine singular object, where the feminine allomorph is -ak- (cf. **10.8**): shakóyɛtha? *he's hitting her*. The nonsingular third person objective construction consists of the NONSINGULAR THIRD PERSON morpheme (-ak-) *they*, followed by the objective morpheme -ɔ(:)-/-ə(:)y-, homonymous with and conditioned in the same way as the second person allomorphs described in **11.2**. It is preceded by the masculine allomorph h-: hakóyɛtha? *he's hitting them*, hakɔyawi:h *he's given it to them*.

11.6. Before the first person singular subject the masculine (object) allomorph is h-: the first person allomorphs are -e-/-ey-, with -e- occurring before a morpheme whose basic allomorph begins with a consonant or i (and the i is lacking), and -ey- occurring before a vowel other than i: heyétha? *I'm hitting him*, he?tyɔ? *I've seated him*, heya:wi:h *I've given it to him*. Before the second person singular subject the masculine allomorph is he-, the second person allomorph -'hs-/-'hse-, conditioned as described in **6.5**, but with -'hs- and -'hse- in place of s- and se- and with -'hs- rather than j- before i or y: hehsa:wi:h *you've given it to him*, hehsí?tyɔ? *you've seated him*. Before the exclusive dual and plural subjects the masculine allomorph is sh- and the exclusive allomorph -a- (cf. **9.3**): shakhníyɛtha? *we* (*excl. du.*) *are hitting him*, shakwáyɛtha? *we* (*excl. pl.*) *are hitting him*. Before the inclusive and second person dual and plural subjects (**9.4–5**) the masculine allomorph is she-: shehníyɛtha? *we* (*incl. du.*) *are hitting him*, shetwáyɛtha? *we* (*incl. pl.*) *are hitting him*, shesníyɛtha? *you* (*du.*) *are hitting him*, sheswáyɛtha? *you* (*pl.*) *are hitting him*.[8] Before the unrestricted third pesron subject the masculine allomorph is h-; the UNRESTRICTED THIRD PERSON morpheme has the allomorphs -ɔwɔ-/-ɔwæ-/-ɔwɛ-/-ɔwɔy-/-ɔw-, with -ɔwɔ- occurring before a morpheme whose basic allomorph begins with a consonant other than (C)æ or (y)ɔ, -ɔwæ- occurring before (C)æ, -ɔwɛ- before (y)ɔ and i (and the i is lacking), -ɔwɔy- before o or ɔ, and -ɔw- before a, e, or ɛ: hɔwóyɛtha? *he, she, they are hitting him*, hɔwé?tyɔ? *he, she, they*

[7] Some idiolects have -awɔ- etc.

[8] The last two forms are homonymous with their reverse, *he . . . you* (*du. or pl.*) (**11.5**).

have seated him, hɔwɔ:wi:h, *he, she, they have given it to him.*

11.7. Another morpheme which occurs only in the transitive prefix is the RESTRICTED THIRD PERSON (ʔ-/ʔe-/-he-/-hey-) *she, they*, excluding the masculine singular meaning but including the masculine nonsingular. This morpheme occurs in a way that is similar but not identical to the occurrence of the masculine morpheme described in **11.4–6**.

11.8. As a subject it precedes the first and second person nonsingular objective construction (**10.11–12**), which, however, occur only with the nonsingular number morpheme (-hi-/-hiy-/-i-/-iy-) rather than with the dual and plural. Its allomorph is ʔ- before the objective morpheme (-ɔ-) that precedes the first person, and ʔe- before the second person morpheme (here -ts-), which conditions a zero allomorph of the objective. The nonsingular allomorphs -hi-/-hiy- occur after the first person morpheme, and -i-/-iy- occur after the second person. The forms without y occur before a morpheme whose basic allomorph begins with a consonant or i (and the i is lacking); the forms with y before a vowel other than i. Examples are ʔɔkhíyɛtha? *she, they are hitting us*, ʔɔkhiyawi:h *she, they have given it to us*, ʔetsíyɛtha? *she, they are hitting you (nonsg.)*, ʔetsiyawi:h *she, they have given it to you (nonsg.)*.

11.9. As an object, unlike the masculine morpheme, the restricted third person FOLLOWS the first (k-) and second (s-) person subjects. In this environment the restricted third person allomorphs are -he-/-hey-, the former occurring before a morpheme whose basic allomorph begins with a consonant or i (and the i is lacking), the latter before a vowel other than i: kheyétha? *I'm hitting her, them*, kheya:wi:h *I've given it to her, them*, sheyétha? *you're hitting her, them*, sheya:wi:h *you've given it to her, them*.

11.10. The restricted third person object does, however, precede its other subjects, the exclusive, inclusive, and second person nonsingular, which (as in **11.8**) occur with the nonsingular morpheme rather than the dual and plural. The nonsingular allomorphs are -hi-/-hiy-/-i-/-iy-, conditioned as in **11.8**, with -hi-/-hiy- occurring also after the inclusive person morpheme. Before the exclusive person nonsingular subject the restricted third person allomorph is ʔ- and the exclusive person allomorph -a-: ʔakhíyɛthaʔ *we (excl.) are hitting her, them*. Before the inclusive person nonsingular subject the restricted third person allomorph is ʔe-; the inclusive person allomorph is -t-: ʔethíyɛthaʔ *we (incl.) are hitting her, them*. Before the second person nonsingular subject the restricted third person allomorph is also ʔe-; the second person allomorph is -ts-: ʔetsíyɛthaʔ *you (nonsg.) are hitting her, them.*[9]

11.11. It may have been noticed that the restricted third person morpheme does not occur as subject with a first or second person singular object. Instead, its meanings are divided between two other morphemes: the feminine (ʔ-/ʔe-),[10] already familiar, and a new one, the NONSINGULAR THIRD PERSON (h-/hɔ-) *they*. The first and second person objective construction follows the subjects, and occurs in the allomorphs already described (**10.6–7**), but the objective allomorph is -ɔ- before the first person. The feminine and nonsingular third person allomorphs are ʔ- and h- respectively before the first person object, ʔe- and hɔ- before the second person: ʔɔkyétha? *she's hitting me*, hɔkyétha? *they're hitting me*, ʔesáyɛthaʔ *she's hitting you (sg.)*, hɔsáyɛthaʔ *they're hitting you (sg.)*.

11.12. There is also a transitive constitute consisting of a neuter subject and neuter object. As the first constituent, the neuter morpheme has the shape k-; as the second, the shape -ɔwɔ-/-ɔwæ-/-ɔwɛ-/-ɔwɔy-/-ɔw-, homonymous with and conditioned in the same way as the unrestricted third person morpheme described in **11.6**: kɔwóyɛthaʔ *it's hitting it*, kɔwɔ:wi:h *it has given it to it.*

11.13. Finally, there are three other members of the transitive construction which

[9] This form is homonymous with its reverse, *she, they . . . you (nonsg.)* (**11.8**).

[10] Some idiolects have ʔɔ- rather than ʔe-.

occur as constituents with a number morpheme (cf. **11.3**), in this case the nonsingular morpheme (-ti-/-n-). The transitive constituents are sh|ako- *he|her* (**11.5**) and two others not yet encountered. These have the masculine (h-) and nonmasculine (k-) morphemes as objects, followed by the restricted third person morpheme (-ɔwɔ-/-ɔwɛ-) as subject. The nonsingular morpheme in all three cases applies semantically to the first of the two constituents, whether it is subject (as with shako-) or object (as with hɔwɔ-/hɔwɛ- and kɔwɔ-/kɔwɛ-). The nonsingular allomorphs -ti- and -n- are conditioned as described in **10.13**. The restricted third person allomorph -ɔwɔ- occurs before -ti-, -ɔwɛ- before -n-. Examples are shakotiyétha? *they (masc.) see her*, shakonɔwi:h *they (masc.) have given it to her*, hɔwɔtiyétha? *she, they are hitting them (masc.)*, hɔwɛnɔwi:h *she, they have given it to them (masc.)*, kɔwɔtiyétha? *she, they are hitting them (nonmasc.)*, kɔwɛnɔwi:h *she, they have given it to them (nonmasc.)*.

11.14. The word-initial indicative allomorph is ?o?- before any transitive prefix beginning with k or s. It is wa- before any other. The optative and future allomorphs are ?aa- and ?ɛ- respectively before any transitive prefix.

11.15. After any modal prefix, any transitive prefix beginning with h in word-initial position followed by a vowel occurs without the h, but with the preceding vowel basically strong: waɔsayɛt *they hit you*, ?áɔsa:yɛt *they might hit you*, ?ɛɔsayɛt *they will hit you*. Also after any modal prefix, the masculine allomorph is -':- rather than -0- before the second person singular object (cf. **11.5**): wa:-ya:yɛt *he hit you*, ?á:yayɛt *he might hit you*, ?ɛ:ya:yɛt *he'll hit you*. After the optative and future morphemes, but not the indicative, any transitive prefix beginning with ? occurs with y in place of the ?: ?á:yɔkhi:yɛt *she, they might hit me*, ?ɛyɔ́khi:yɛt *she, they will hit me*. With these exceptions, the same allomorphs occur after the modal prefix morphemes as in word-initial position.

11.16. The following is a list of the transitive prefixes:

kɔ(:)-/kə(:)y- *I . . . you*: kɔ:kɛh *I see you*, ?o?kɔ:kɛ? *I saw you*, ?ɛkɔ:kɛ? *I'll see you*.

sk-/ske-/tak-/take- *you . . . me*: ske:kɛh *you see me*, ?o?ske:kɛ? *you saw me*, ?ɛske:kɛ? *you'll see me*.

khni-/khni/khny- *I, we . . . you (du.)* (**11.3**): khni:kɛh *I, see we you*, ?o?khni:kɛ? *I, we saw you*, ?ɛkhni:kɛ? *I, we will see you*.

kwa-/kwæ-/kwɛ-/kw-/ky- *I, we . . . you (pl.)*: kwa:kɛh *I, we see you*, ?o?kwa:kɛ? *I, we saw you*, ?ɛkwa:kɛ? *I, we will see you*.

skhni-/skhn-/skhny- *you . . . me, us (du.)*: skhni:kɛh *you see me, us*, ?o?skhni:kɛ? *you saw me, us*, ?ɛskhni:kɛ? *you'll see me, us*.

skwa-/skwæ-/skwɛ-/skw-/sky- *you . . . me, us (pl.)*: skwa:kɛh *you see me, us*, ?o?skwa:kɛ? *you saw me, us*, ?ɛskwa:kɛ? *you'll see me, us*.

hak-/hake- (/-'ak- etc.) *he . . . me*: hake:-kɛh *he sees me*, wa:ke:kɛ? *he saw me*, ?ɛake:-kɛ? *he'll see me*.

ya-/yæ-/yɛ–/y- (/-':ya- etc.) *he . . . you*: ya:kɛh *he sees you*, wa:ya:kɛ? *he saw you*, ?ɛ:ya:kɛ? *he'll see you*.

shɔkhni-/shɔkhn-/shɔkhny- *he . . . us (du.)*: shɔkhni:kɛh *he sees us*, ?o?shɔ́khni:kɛ? *he saw us*, ?ɛshɔ́khni:kɛ? *he'll see us*.

shɔkwa-/shɔkwæ-/shɔkwɛ-/shɔkw-/shɔky- *he . . . us (pl.)*: shɔkwa:kɛh *he sees us*, ?o?shɔkwakɛ? *he saw us*, ?ɛshɔkwakɛ? *he'll see us*.

shesni-/shesn-/shesny- *he . . . you (du.)*: shesni:kɛh *he sees you*, ?o?shéshni:kɛ? *he saw you*, ?ɛshéshni:kɛ? *he'll see you*.

sheswa-/sheswæ-/sheswɛ-/shesw-/shej- *he . . . you (pl.)*: sheswa:kɛh *he sess you*, ?o?shéswakɛ? *he saw you*, ?ɛshéswakɛ? *he'll see you*.

shako-/shaka-/shakaw- *he . . . her*: shako:-kɛh *he sees her*, ?o?shako:kɛ? *he saw her*, ?ɛshako:kɛ? *he'll see her*.

hakɔ(:)-/hakɔ(:)y- (/-'akɔ[:]- etc.) *he . . . them*: hakɔ:kɛh *he sees them*, wa:kɔ:kɛ? *he saw them*, ?ɛɔkɔ:kɛ? *he'll see them*.

he-/hey- (/-'e- etc.) *I . . . him*: he:kɛh *I*

see him, wa:ekɛʔ *I saw him*, ʔɛ:ekɛʔ or ʔe:ekɛʔ *I'll see him* (**3.16**).

héhs-/héhse- (/-'éhs- etc.) *you . . . him*: hehse:kɛh *you see him*, wáehse:kɛʔ *you saw him*, ʔé:hse:kɛʔ *you'll see him*.

shakhni-/shakhn-/shakhny- *we (excl. du.) . . . him*: shakhni:kɛh *we see him*, ʔoʔshákhni:kɛʔ *we saw him*, ʔɛshákhni:kɛʔ *we'll see him*.

shakwa - /shakwæ - /shakwɛ - /shakw - /shaky- *we (excl. pl.) . . . him*: shakwa:kɛh *we see him*, ʔoʔshakwakɛʔ *we saw him*, ʔɛshakwakɛʔ *we'll see him*.

shehni-/shehn-/shehny- *we (incl. du.) . . . him*: shehni:kɛh *we see him*, ʔoʔshéhni:kɛʔ *we saw him*, ʔɛshéhni:kɛʔ *we'll see him*.

shetwa - /shetwæ - /shetwɛ - /shetw - /shety- *we (incl. pl.) . . . him*: shetwa:kɛh *we see him*, ʔoʔshetwakɛʔ *we saw him*, ʔɛshetwakɛʔ *we'll see him*.

shesni-/shesn-/shesny- *you (du.) . . . him*: shesni:kɛh *you see him*, ʔoʔshésni:kɛʔ *you saw him*, ʔɛshésni:kɛʔ *you'll see him*.

sheswa- / sheswæ- / sheswɛ- / shesw- / shej- *you (pl.) . . . him*: sheswa:kɛh *you see him*, ʔoʔshéswakɛʔ *you saw him*, ʔɛshéswakɛʔ *you'll see him*.

hɔwɔ- / hɔwæ- / hɔwɛ- / hɔwɔy- / hɔw-(/-'ɔwɔ- etc.) *he, she, they . . . him*: hɔwɔ:kɛh *he, she, they see him*, waɔwɔkɛʔ *he, she, they saw him*, ʔɛɔwɔkɛʔ *he, she, they will see him*.

ʔɔkhi-/ʔɔkhiy- (/-yɔkhi- etc.) *she, they . . . us*: ʔɔkhi:kɛh *she, they see us*, waʔɔ́khi:kɛʔ *she, they saw us*, ʔɛyɔ́khi:kɛʔ *she, they will see us*.

ʔetsi-/ʔetsiy- (/-yetsi- etc.) *she, they . . . you (nonsg.)*: ʔetsi:kɛh *she, they see you*, waʔétsi:kɛʔ *she, they saw you*, ʔɛyétsi:kɛʔ *she, they will see you*.

khe-/khey- *I . . . her, them*: khe:kɛh *I see her, them*, ʔoʔkhe:kɛʔ *I saw her, them*, ʔɛkhe:kɛʔ *I'll see her, them*.

she-/shey- *you . . . her, them*: she:kɛh *you see her, them*, ʔoʔshe:kɛʔ *you saw her, them*, ʔɛshe:kɛʔ *you'll see her, them*.

ʔakhi-/ʔakhiy- (/-yakhi- etc.) *we (excl.) . . . her, them*: ʔakhi:kɛh *we see her, them*, waʔákhi:kɛʔ *we saw her, them*, ʔɛyákhi:kɛʔ *we'll see her, them*.

ʔethi-/ʔethiy- (/-yethi- etc.) *we (incl.) . . . her, them*: ʔethi:kɛh *we see her, them*, waʔéthi:kɛʔ *we saw her, them*, ʔɛyéthi:kɛʔ *we'll see her, them*.

ʔetsi-/ʔetsiy- (/-yetsi- etc.) *you (nonsg.) . . . her, them*: ʔetsi:kɛh *you see her, them*, waʔétsi:kɛʔ *you saw her, them*, ʔɛyétsi:kɛʔ *you'll see her, them*.

ʔɔk-/ʔɔke- (/-yɔk- etc.) *she . . . me*: ʔɔke:kɛh *she sees me*, waʔɔke:kɛʔ *she saw me*, ʔɛyɔke:kɛʔ *she'll see me*.

hɔk-/hɔke- (/-'ɔk- etc.) *they . . . me*: hɔke:kɛh *they see me*, waɔke:kɛʔ *they saw me*, ʔɛɔke:kɛʔ *they'll see me*.

ʔesa-/ʔesæ-/ʔesɛ-/ʔes- (/-yesa- etc.) *she . . . you*: ʔesa:kɛh *she sees you*, waʔesakɛʔ *she saw you*, ʔɛyesakɛʔ *she'll see you*.

hɔsa-/hɔsæ-/hɔsɛ-/hɔs- (/-'ɔsa- etc.) *they . . . you*: hɔsa:kɛh *they see you*, waɔsakɛʔ *they saw you*, ʔɛɔsakɛʔ *they'll see you*.

kɔwɔ-/kɔwæ-/kɔwɛ-/kɔwɔy-/kɔw- *it . . . it*: kɔwɔ:kɛh *it sees it*, ʔoʔkɔwɔkɛʔ *it saw it*, ʔɛkɔwɔkɛʔ *it will see it*.

shakoti-/shakon- *they (masc.) . . . her*: shakoti:kɛh *they see her*, ʔoʔshakoti:kɛʔ *they saw her*, ʔɛshakoti:kɛʔ *they'll see her*.

hɔwɔti-/hɔwɛn- (/-'ɔwɔti- etc.) *she, they . . . them (masc.)*: hɔwɔti:kɛh *she, they see them*, waɔwɔti:kɛʔ *she, they saw them*, ʔɛɔwɔti:kɛʔ *she, they will see them*.

kɔwɔti-/kɔwɛn- *she, they . . . them (nonmasc.)*: kɔwɔti:kɛh *she, they see them*, ʔoʔkɔwɔti:kɛʔ *she, they saw them*, ʔɛkɔwɔti:kɛʔ *she, they will see them*.

NATICK CONSONANTS IN REFERENCE TO PROTO-CENTRAL ALGONQUIAN: II*

SHIRLEY SILVER

UNIVERSITY OF CALIFORNIA, BERKELEY

2.2. List of cognates, (Cont.), M-Y
3. Single consonant correspondences
4. The 'intrusive' nasal
5. Consonant cluster correspondences

2.2. MALE. PCA *na·pe·w-a (Bl); Nt. nampvs (nompaas *male of the human species*).

MAN. PCA *elenyiwa (Bl) †elenyiwaki, PL.; Nt /ninwak/ (ninnuog *men*). It is not clear from Trumbull's dictionary whether this is a Natick or Narragansett form (ninnuog) is listed in the English-Natick entries, but in the Natick-English part we find only (nnìn) *man* which is starred, a device used to indicate that the form 'does not belong to the Natick dialect proper'. Under this starred form, which is Narragansett, is listed a plural (nnínnuog, nínnuock).

MEDICINE. See ROOT.

MEET. PCA *naθkwe·škawe·wa (*naθkw- + -e·- + -ešk-aw- *by foot/body movement*) *he goes so as to correspond with the other* (Bl); Nt nakəškaw (nogkushkau, nogkishkau *he meets him*).

MIXED TOGETHER. PCA *kelek- (H); Nt kinək- (kenukshau *it is mingled or mixed with*, nuk-kinukkinum *I mingle or mix*).

MOUTH: HIS ---. PCA *weto·n-i (M); Nt /wətun/ (wuttoon).

MOVE. PCA *ma·t-, *ma·cyi·hkani (F ma·či·hkani *thous mightest move*, C ma·či·hkan *do thou later go on a hunt*, O ma·či·kkan *do thou later go away*) (Bl); Nt /mɛmantənəm/ (mamontunnum *he moves/puts it in motion*), /mɛmančit/ (mamonchit *when he moves*, cf. monchanaü *he moves him*). Cp. Nr maûchase *Be my guide!*.

MUCH. See BIG (2).

NAIL. PCA *-škanš-a (Ha) (K.Ha neškaša, C.L n'iskasïy, O.Ba nishkânj *my nail*, Sh -kaše, -ʔškaša); PCEA *-xkanš-a (Ha) (D nhíkaš *my nail*); Nt /məʰkas/ (múhkos, múhkas *the nail of a man, the hoof, claw talon of an animal*, cf. wuhkossog, wuhkassog *his nails*).

NAME (*vb*). PCA *wi·nle·wa (Bl) *he names him*, *wi·ntamwa (Bl) *he names it, tells it*; Nt /əsuwinaw/ (ussowenau *he names him*), /əsuwitɛm/ (ussowetam *he calls or names it*), /əsuwitɛmənvt/ (ussoowetamunát *to give a name to inanimate object*).

NEGATIVE. PCA *kaθ- (or *kan-) (H);[19] Nt /-vʰkan/ the 2nd sg. of the imperative negative is formed in (-âhkon or -uhkon), cf. (kummootuhkon) *thou shalt not steal*.

NET. PCA *aʔlapya (Bl); Nt /(h)ešep/ (hasháb, hashábp *a net for fish*, âshâp, âshâppog *vegetal fiber or fibrous material used for making thread or cord*). For correspondence of PCA *ʔl: Nt/š/cp. FLY.

NEW. PCA *weški (Bl); Nt /wəski/ (wuske, weske *young, new*).

OLD PERSON. PCA *keʔtesiwa (H); Nt /keʰčisw-it/ (kuhchisu-it *when he is old*, cf. kehchisqua, kutchisqua *an old woman*, kehchiog, kutchiog *the old* [collectively]).

ONE. PCA *nekot- (Bl); Nt /nikwət/ (nequt).

OTHER. PCA *kotak- (H); Nt /anketak/ (onkatog *another person*, cf. onkatogig *others*).

PLACE: HE ---S HIM. PCA *aʔle·wa (Bl); Nt /panaw/ (ponau). This set is to be questioned because of the unexplained initial consonant in Nt. For discussion of the prob-

[19] Hockett; ibid, "M has the negator with the shapes /kan/ and /kat/; it is this alternation that points to *kaθ- rather than *kan-."

* For Part I see IJAL 26.112–119 (1960).

lematical correspondence of Nt /n/ with PCA *?l see FLY.

PLAY. PCA *pa·hp-, *pa·hpiwa *he laughs* (Bl); Nt /pampu/ (pompu, pohmpu *he plays*, cf. pompuog, pohmpuog *they play*, puhpŭog *they play*).

PLURAL, ANIMATE. PCA *-aki (Bl); Nt /-ak/ (-og), cf. SNAKE.

POT. PCA *aθkehkw-a (Bl) *kettle*; Nt /aʰkəʰk, aʰkəʰkw-/ (ohkuk, ohkuhk, ahkuhq *earthen pot or vessel*, ahkuhquog *pl.*).

POUND. PCA, PCEA *takw- (?; Ha) (C.L takkwahwew *piler*, Sh takhw- *pound, grind for bread*, Mi tähkwahikwani *I grind it in a mill*, Mc tekwtāāgā *to grind grain*); Nt /takəhusu/ (togkuhwhosu, togguhwhosu *he grinds in a mill*).

PRICKLY. PCA *ka·w- (H); Nt /kuwa/ (koo, koowa, koowâs *pine tree*, cf. kôus *a thorn, briar, bramble* [with dimin. suffix ?]). Cp. Pn.Ha kò·wi *quill* (*porcupine*).

PRONOMINAL BASE. PCA *-i·l- (Bl); Nt /-in/, cf. PCA *ni·la; Nt /nin/ (neen, nen) *pers. pron., 1st sg.* PCA *ki·la *thou*, *ki·lwa·wa *ye*; Nt /kinaw, kinawaw/ (kenau, kenauau *you*). PCA *ni·la·na; Nt /ninawən/ (neenawun) *pers. pron. we* (*excl.*).

PRONOMINAL PREFIX (FIRST PERSON). PCA *ne- (Bl); Nt /nə-/ (nu-), cf. RELATED, SISTER (1).

PRONOMINAL PREFIX (SECOND PERSON). PCA *ke- (Bl); Nt /kə-/ (ku-), e.g. *kete·hi (Bl); Nt /kəte(·)/ (kuttah) *thy heart.*

PRONOMINAL PREFIX (THIRD PERSON). PCA *we-, *w- (Bl); Nt /wə-, w-/ (we-, w-), cf. HEART, LEG, TONGUE, TOOTH.

PRONOMINAL PREFIX (THIRD PERSON, INDETERMINATE). PCA *me- (Bl); Nt /mə-/ (mu-), cf. THROAT, KNEE.

PUT INTO. PCA *pi·nt- (Ha) *within*; Nt /pitaw/ (pētau *he puts it into*).

RED. PCA *meskw- (Bl); Nt /məˢ| škwi/ (musqui, mishque, msqui *it is red*).

RELATED. PCA *neteθa·nko·ma·waki (G) *I have them as relatives;* Nt /nətinankəmowak/ (nuttinonkumóog *my kinsman*). Cp. Sh nitela·ko·ma·ki *I am related to them*, Dl ntəlankú·ma *I'm related to him.*

ROCK. PCA, PCEA *-a·peθk-w- (Bl) *stone, metal;* Nt /-pisk/ (-pisk *rock in compound words*, cf. chippisk, chepisk *rock, crag*), /-ampskw/ (-ompsk, ompsq *rock in compound words*, cf. kenompsq *sharp stone*).

ROLL (*vb.*). PCA, PCEA *te·tep- (Ha); Nt /tətəp . . . / (tatuppequanum *he rolls it along*, tattuppunau *it is twisted*, tuttuppun *spun, twisted*).

ROOT, MEDICINE. PCA, PCEA *we-ce·pihk- (?; Ha); Nt /wəčapəʰk/ (wadchâbuk, wutchaubuk, -adchábuk [in compounds], wutchappehk *root*). The *hk is questioned by Haas since the Cree form shows no evidence for it: C.W ochapik, eO uci·pikk, Sh hoče·pi?k-i. It is likewise questionable for PCEA since, even though the Natick form may be evidence for an *hk, the Mc oochebusk points to *θk.

SCRAPE. PCA *ka·šk- (H); Nt /kəški/ (kushki *it is rough*). This is a likely set; however, cf. Haas' discussion of the difficulties with Hockett's reconstructions for Scrape and Dry.[20] Cf. also F.H kahkahkya·wi *it is rough* (included in the set under Hockett's reconstruction *kaHk- (*H: *h or *θ) *sharp* and Nt (koshkeyĕue) *roughly* (*a coat not soft*).

SEA. PCA *ke?čikami (H); Nt (kehtohhan, kehtahhan *any sea*, kehtoh, keihtoh *the ocean*). Cp. Nr kitthan, wechêkum *the sea.* SEE BIG (2).

SEE: HE ---S HIM. PCA *ne·we·wa (Bl); Nt /nvw/ (nâaw).

SEED. PCA *-θka·n (Ha), *mi·n- (Ha), *mi·nali (Bl) *berries*; Nt /skənimənəš/ (skannémunash *seeds, seed corn*), /-mini/ (-minne *berry in compound names*, cf. minneash *small fruit of any kind*).

SEIZE. See BITE.

SHAKE. PCA *neki- (?; Ha) (Sh naneky-, neneki, *shaking nervously*); Nt /nəkimu, nənəkimu/ (nukkemoo, nunnukkimoo *it*

[20] Haas, Mary R.; Notes on Some PCA Stems in /k-/, IJAL 24.241–245 (1958).

shakes, it is shaken). Natick, like Shawnee, shows reduplication.

SHARP. PCA *ki·n-, *ki·n-t (Ha), *ki·n-l (H); Nt /kiʰtɛtaw/ (kéhtadtau, kehtattau *he sharpens*), /kin-, kini-/ (kenai, kēneh *it is sharp*, kenompsk *a sharp stone*, keneh quog *a sharp knife*).

SHOE. PCA *maθkesini (Bl); PCEA maxkesini (S); Nt /makəs, makəsineš/ (mokus, mokis *shoe*, mokussinash *shoes*).

SHOOT. PCA *pemwe·wa (Bl) *he shoots him*; Nt /pipəmwaw/ (pepumwau *he shoots at him*). The Nt form shows reduplication.

SHORT. PCA *tahkw- (Ha), †tahkwisiwa *he is short;* Nt /tiaʰkwsiw/ (tiohqoosue *he is short*, cf. tióqui *it is short*). Cp. Nr tiaquónqussu *he is short.*

SISTER (1). PCA *nemihsa (Bl) *my elder sister;* Nt /nəmisis/ (nummissis *my sister*).

SISTER (2). PCA *-i·tekeHkw- (Bl; *H: ?) *sister* (*woman speaking*), †ni·tekeHkwa *my sister*, †wi·tekeHkwa *her sister;* Nt /nitəkaˢ|škw/ (netukkusq *my sister*), /witək(i)ˢ|škwa/ (wetukishquoh, weetuksquoh *her sister*).

SIT: He ---s. PCA *apiwa (Bl), †apiwaki *they sit*; Nt /ɛpw/ (appu *he sits*), /ɛpwak/ (appuog *they sit*).

SLEEP WITH. PCA *wi·θpe·me·wa (Bl), PCEA *wi·xpe·me·wa *he sleeps with him*; Nt /weʰpɛmaw/ (wehpamau *he has sexual connection with her*).

SNAKE. PCA, PCEA *aθko·ka, -aki pl. (S); Nt /ɛskuk, ɛskukak/ (askook, askookog *pl.*).

SNOW (*n.*). PCA *ko·n (H); Nt /kun/ (koon).

SNOW: IT ---s. PCA *meθponw-i, PCEA mexponw-i (S); Nt /məʰpuwi/ (muhpoowi).

SOFTEN. PCA *lo·θk-ah- (G) *soften by pounding*; Nt /nuʰki/ (noohki *adj. soft*, cf. noohkik *that which is made soft, i.e. meal, flour, ground corn*). Cp. Nr nókehick *parched meal*, Pn.Ha nóhkikən *it's soft.*

SOILED. See EXCREMENT.

SO MANY. PCA *tahθ-, *tahšiwaki *they* (*an.*) *are so many* (Bl); Nt /taʰsv, taʰšiw/ (tohsú, tohshéu *how much*), /taʰswak/ (tohsuog *how many* (*an.*), cf. tohsunash *how many* (*inan.*), tohshe *how much*). Notice the recording in Natick of wha would seem to be /š/ before /i/ and /s/ elsewhere. If this alternation has an reality, a problem to be solved by a mor detailed synchronic examination of th Natick corpus available, we have here a example of Natick reflecting the PC morphophonemic alternation between *(C) and *(C)š.

SOUTH: IN THE ---. PCA *ša·wanoni (Ha); Nt /sawɛnaʰki (sowanohke *the south country*).

SPIT. PCA *sekw-, *sehkw-i (?; Ha) Nt /səʰk ..., səʰkw/ (suhkou *he spit* sookq, sohq *spittle*, suhkesu *he is a spitte* sukkissu *long clam, spitter*). Cp. Sh θekw *spit*, Cr(W) si'koo *spit* (*v.i.*).

SPOTTED. PCA *mya·l- (Bl); Nt /m moni/ (mômóne *freckled*). The Natick for shows reduplication.

SPROUT (*vb*). PCA *sa·kihθenwi (M *it sprouts*; Nt /sanki ... / (sonkin *it spring up* [*as a plant*]).

STAND. PCA *-ka·pawi- (Bl); Nt /kan paw/ (kompau *he stands erect* [occurs onl in compounds], cf. sampookompau *he stand upright*).

STAR. PCA *aθa·nkwa (Bl); Nt /ɛnakws (anogqs). The final /s/ in Natick is prob ably the diminutive suffix.

STEAL. PCA *kemo·twa (B1); *he steals* Nt /kamutv/ (kommooto, kummooto *steals*).

STONE. PCA *a?senyali (Bl) *stone* Nt /kwineʰsinənk (quinahsinnunk *a pestl i.e. long stone* < qunni, quinne *long*, hassu assin *stone*).

STRONG. PCA *sa·pw- (G) *strong;* /sampwi/ (sampwi *straight, right, just* Cp. Nr saûmpi *straight*, Qu.T sompâ *it is right.* This set is questionable because the possibly tenuous semantic relationship

STRETCH. PCA *ši·pi (Ha); Nt /sip (sepe *it spreads out, extends*).

SUCKLE. PCA *no·n- (Bl); Nt /nənu

(nunoon *I suck*), /nunək/ (noonuk *a sucking*).

SUMMER. PCA ni·pin (Ha); Nt /nipən/ (népun *the latter part of the summer*).

SWEET. PCA *wi·nk-an- (G); Nt /wikan/ (weekon *it is sweet*). Cp. Sh wiikanwi *it tastes good.*

TASTE. PCA, PCEA *eθpokw- (Ha); Nt/spəʰkwat/ (spukquodt, spuhquodt *the taste, flavor of a thing*). Cp. Nr teàqua aspùckquat *what does it taste of.*

TEACH. PCA *keθkinw-, *keθkino·hama we·wa (H) *he teaches it to him (trans. an.)*; Nt /kəʰkin . . . / (kuhkinneam *he takes note of [mentally or by observation, v.t.]*, cf. kuhkootumwehteau *he teaches [caus. inan.]*).

TELL NEWS: HE ---S. PCA *a·cimoke·wa (G); Nt /ančimukaw/ (aunchemookau). Cp. Sh haačimo *he talks*, haat- (plus -i-m-, -ota) *talking*, Mi ačimokiani *I boast.*

THICK: IT IS ---. PCA *keXpakya·wi (H); Nt /kəpaki/ (kuppogki). Cp. Sh kpakinwi *it is thick* (< pre-Sh *kV?pakinwi).

THINK. PCA *-e·le- (Bl), †eθe·letamwa *he thinks so* (O ine·ntam); Nt /ɛnɛntɛm/ (anántam *he thinks*, cf. nuttenantam *I think*).

THREE. PCA *ne?θ-, *ne?θwi (Bl); Nt /niš/ (nish *three*), niswi (nishwe, nishweu *the third*).

THROAT: SOMEONE'S ---. PCA *mekonta·kani (M); *ko(·)nta·kani (H) *throat;* Nt /məkwətənk/ (mikquttunk), /kəkwətənkɛnit/ (kuk-quttunkanit *to thy throat*). Cp. Nr gùttuck *throat.*

THUMB. PCA *ke?ciθencyi (H); Nt /keʰtvkwɛnič/ (kehtequanitch [keht *great*, uhquae *end of*, -nutch *hand*], kehtequanitcheash *pl.*). SEE BIG (2), FINGER.

TOE: BIG ---. PCA *ke?cisita (H); Nt /keʰtvkwɛsit/ (kehtequaseet: keht *great*, uhquae *end of*, -seet *foot*). See BIG (2) FOOT.

TONGUE: HIS ---. PCA *we·θan-i (M); Nt /winɛn/ (wēnan, weenan).

TOOTH: HIS ---. PCA *wi·pit- (Bl); Nt /wipit/ (weepit).

TOP, ON. PCA, PCEA *waθkici (S) *on top, on the surface*; Nt /waskiči/ (woskeche *on the surface of*).

TREE. PCA *me?tekw- (Bl); Nt /meʰtəkw-/ (mehtug, mehtugq *tree*, mehtugquash, matugquash *trees*); PCA *-a·htekw- (Bl); Nt /wəhtəʰkw/ (wuttuhq, wuttuk, wuhtuk, *bough, branch*, cf. wuttoohqŭnash *wood pl.*).

TRY. PCA *kot-, *koci (*part.*) *as preverb try to*, *kočihe·wa *he tries, tests him*, *koci?ta·wa *he tries, tests it* (H); Nt /kwəčihaw/ (qutchehheau *he tempts, proves him*), kwəčiʰtaw (qutchehteau *he proves it*).

TWO. PCA *ni·š-, ni·šwi (Bl); Nt /nis, niswi/ (neese, neeswe).

UNCOVER: HE ---S IT. PCA. PCEA *pa·θkenamwa (S); Nt /paskinəm/ (poskinum *he bares it*).

URINATE. PCA *šeki- (M); Nt /sɛkitak/ (noh ságketog, noh sagkeet *he who urinates*).

VIEW (*vb.*). PCA *wensa·pame·wa (Bl) *he sees him from there*; Nt /wəsɛmpvnɛt/ (wussampenat *to view, look out from*).

WALK BY. PCA *pemohθe·wa (M) *he walks by*; Nt /pamšaw/, (pomsheau, pomshau *it passes, goes by*, cf. noh pomushadt, pamwushadt *he who walks*).

WATCH (*vb.*). PCA, PCEA *aθk- (S) *watch, lie in ambush*; Nt /ɛskəhəm/ (askuh hum *he watches, waits for it*).

WATER. PCA *nepyi (Bl); Nt /nəpi/ (nippe, nuppe).

WHITE. PCA *wa·p- (Bl); Nt /wampi/ (wompi *there is white*).

WHO. PCA *awe·n- (Bl); Nt /hawɛn/ (howan).

WIDOW. PCA, PCEA *si·kawe θkwe·wa (S); Nt /sikawˢ|škw/ (sekousq, sekauishq, sekousquaog *pl.*). See WOMAN.

WINTER. PCA, PCEA *pepo·nw- (Ha); Nt /pvpvn/ (popón, pópon *it is winter*).

WITH. PCA *wi·t-, *wi·č- (Bl); Nt /witamaw/ (weetomau *she dwelt with her*), wiči (weeche *with, in company with person or animate object*).

WOMAN. PCA, PCEA *eθkwe·wa, (Bl); Nt /(e)ˢ|škw, ɛskwɛs/ (squa, eshqua *female*,

squáas, squaus *a woman with termination denoting a living creature*).

WOODCHUCK. PCA, PCEA *akwaθkwa (S); Nt /akaškwak/ (ogkoshquog *conies*).

WORK: HE ---S. PCA *aθoθkye·wa, PCEA *aθoxkye·wa (S); Nt /ɛnɛkawsu/ (anakausu, cf. anakausish *labor thou!*).

YELLOW. PCA †wesa·wa (C.L osâwaw, O osâwa *it is yellow*, Sh hoθaaw *it is yellow*); Nt /wisowi/ (weesóe *yellow*). See BITTER.

YESTERDAY. PCA *weθa·kw- (Ha); Nt /wənanku/ (wunnonkou, wunnonkoo *it was evening, yesterday, last evening*). Cp. Mc oolagoó *evening*, Mi alakiä *yesterday*, Nr wunnáuquit *evening*.

3. Examples of the PCA--Natick single consonant and semivowel correspondences are cited below according to English gloss. See Chart II for a comparison of the PCA reconstructed phonemes and their F, C, M, O and Sh reflexes with the Natick correspondences.[21]

CHART II

F	C	M	O	Sh	PCA	Nt
n	t, c	n	n	l	*θ	n
n	y	n	n	l	*l	n
n	n	n	n	n	*n	n
y	y	y	y	y	*y	y
t, c	t, c	t, c	t, c	t, c	*t, *c	t, č
p	p	p	p	p	*p	p
k	k	k	k	k	*k	k
c	c	c	c	c	*c	č
s	s	s	s	θ	*s	s|š
š	s	s	š	š	*š	s|š
h	h	h	h	h, ?h	*h	h
w	w	w	w	w	*w	w
m	m	m	m	m	*m	m

*θ: n. *abandon, again, armpit, between, breast, dog, fight, finger, negative, related, star, think, tongue, work, yesterday.*

[21] Although the PCA reconstructions were made on the basis of four Central languages only, i.e. F, C, M, and O, the Sh reflexes fit the reconstructions; cf. Wick R. Miller, op. cit.

*l: n. *bark, breathe, dig, dish, five, man, mixed together, pronominal base, soften, spotted, think.*

*n: n. *abandon, arm* (see also *bone, jaw, liver, mouth, throat, tongue*), *bend* (see also *catch hold of, uncover*), *breast, by hand,* (see also *lift up*), *child, daughter, die, fish, go after, hear, hold, kill, lick, long, lose, male, man, meet, one* (see also *eight, two, three*), *pron. prefix* (*first pers.*), *see, seed, shake, shoe, snow* (*n.*), *south, stone, suckle, summer, sweet, teach, water, who, winter.*

*y: y. *body, come, four.*

*t: t. *arm, brain, bring, daughter, day, enough, feet, fire, freeze, heart* (see also *jaw, mouth, related*), *hold, hot, kick, knee, lick, one* (see also *other*), *pound, roll, short, sister* (*2*), *so many, steal, think, tooth.*

*t, *c: t, č. *bad, eat, try, with.* (See Sec. 5.)

*?t, ?c: ʰt, ʰč). There is one example of *t: ʰt (cf. *deer*); however, the /ʰ/ may represent an attempt to indicate the quality or quantity of the preceding vowel.

*p: p. *along in space or time* (see also *fly, live, shoot, walk by*), *bite, break* (*2*), *bring, burst, closed, come, die, dust, enough, enter, five, flea, foam, grease, net, play, pull, into, roll, root, sit, stretch, summer, tooth, uncover, water, winter.*

*k: k. *abandon, big* (*2*), *bite* (see also *catch hold of*), *bitter, closed, day, dip out/up, ear, eight, enter, eye, fear, fight, flea, forehead, freeze, hot, house, knee, long mixed together, negative, one, other, plural* (*an.*), *pound, prickly, pron. prefix* (*second pers.*), *scrape, sharp, sister* (*2*), *snake, snow* (*n.*), *stand, steal, teach, tell news, thick, throat, try, urinate, widow, woodchuck.*

There are three examples of *k: ʰk. Again, this may not be a correspondence but may reflect orthographic conventions; see *break* (*1*), *taste, tree.*

*c: č. *child, root, top.*

*s: s. *bite* (see also *catch hold of*), *eat, feel, go after, shoe, spit, sprout, strong, widow, yellow.*

*š: s. *break (1), day, eight, hot, south, stretch, two, urinate.*

*h: h. *closed* (see also *dip out/up*).

*w: w. *abandon, bend, between, bitter, body, break (1), dig, dish, eat, egg, eight, excrement, foam, four, hot, look, lose, new, prickly, pron. prefix (third pers.), red, root, short, sit, south, spit, stand, sweet, taste, top, tree, two, walk by, white, who, with, woman, yesterday.*

*m: m. *along in space or time* (see also *fly, live, shoot, walk by*), *bad, bear, bee, bend* (see also *catch hold of, closed, dip out/up, fear, find, go after, kick, lick, lift, look, uncover*), *big (1), dog, eat, excrement, fight, fish, grandmother, grease, hair, move, pron. prefix (third pers. indeterminate), red, related, seed, shoe, sister (1), sleep with, spotted, steal, tell news, tree.*

4. In his article on "Algonquian nasaump and napōpi" (Lg. 21.40–45), Geary points out that in the New England languages an original Proto-Algonquian nasal cluster usually loses its nasal and that those nasal clusters which appear are due to an 'intrusive nasal' after ō or au < PA *a·. He suggests that the transcriptions by Eliot and others may actually represent a nasalized vowel plus consonant. The consonant involved is always a stop (or the affricate č).

Examples of PCA *a·T (T: stop), *a·hT corresponding to Nt vowel plus nasal-stop cluster are given below. The most frequent recording of the vowel is (o). There is no conclusive evidence that can be used to determine its precise phonetic value, but for the time being I have normalized this vowel as /a/; cf. English pond.[22] This normalization is presented as a working hypothesis subject to change. No assumptions have been made about the length of the vowel.

*a·p: amp. *again, dip out/up, look, male, rock, stand, strong, view, white.*

*a·hp: amp. *jaw, play.*

*a·t: ant. *leg, live.*

*a·t(*a·c): ant (anč); *move.*

*a·k: ank. *dish, sprout, throat, yesterday.*

: ⁿ/onk. *bend.*

There are exceptions to this correspondence (cf. *break (1)* [Haas questions the PCA form], *freeze, hide, lick*). Because of lack of knowledge of the possible conditioning factors which may be in operation and because of the inconsistencies in transcription, these exceptions cannot be explained. However, if we look at Natick and Narragansett we see that, insofar as one can tell from the orthography, they do not always agree in their treatment of the development of the nasal, e.g. Nt wosketomp, Nr eneskéetomp *man* in contrast with Nt wunnonk, Nr wunnáug *dish*; Nt mukquttunk, Nr qúttuck *throat.*

This 'intrusive' nasal may be a secondary development in Nt and some other New England languages. If it is not, then it is a reflex of a PA (as opposed to PCA) feature.

In order to solve the problem, a systematic comparative study of both the Eastern and Western languages is needed. It may not be entirely out of place to note, e.g. that Cheyenne appears to have hT corresponding to the 'intrusive' nasal + stop, cf. PCA *wa·pi, Nt (wompi), Ch vuhpi *white*; PCA *wela·kani, Nt (wunnonk), Ch hi tuhkU *dish.*

5. The PCA reconstructed consonant clusters with their F, C, M, O and Sh reflexes and the Natick correspondences are presented in Chart III according to the first member of the cluster.[23]

[22] In the material I have examined thus far, wherever a PCA form contains *o and *o· the Nt recording is (oo); cf. Eliot's observation: "Especially we have more frequent use of o and oo than other languages have and our oo doth always sound as it doth in these English words, moody, book.".

[23] The starred forms to the right of the Natick correspondences represent PCEA consonant clusters.

CHART III

F	C	M	O	Sh	PCA	Nt
p	hp	hp	mp	p	*mp	p
t	ht	ht	nt	t	*nt	t, ʰt
k	hk	hk	nk	k	*nk	k, ʰk
c	hc	hc	nc	c	*nc	č
s	s	hs	ns		*ns	s
š	s	hs	nš	š	*nš	s
	ht	hn	n		*nθ	
n	hy	hn	n	n	*nl	n
	hp	hp	pp		*hp	
ht	ht	ht	tt	ʔt	*ht	t, ʰt
hk	hk	hk	kk	ʔk	*hk	k, ʰk
hc	hc	hc	cc	ʔc	*hc	
s	s	hs	ss	ʔθ	*hs	s
š	s	hs	šš	ʔš	*hš	
s	ht	hn	ss	ʔθ	*hθ	š
s	hy	hn	ss	ʔθ	*hl	š
p	sp	hp	pp	ʔp	*Xp	p
hp	sp	hp	pp	ʔp	*θp	p, ʰp → *xp; sp → *θp
hk	sk	hk	kk	ʔk	*θk	k, ʰk → *xk; sk, šk → *θk
	sp	cp	pp		*cp	
hk	sk	ck	šk	ʔk	*ck	
hp	sp	sp	šp	šp	*šp	šp
	st		št		*št	
šk	šk	sk	šk	šk	*šk	sk, šk
						ʰk
šk	hk	hk	sk	šk	*sk	sk, šk
F	C	M	O	Sh	PCA	Nt
ht	st	ʔt	tt	ʔt	*ʔt	ʰt
hc	sc	ʔc	cc	ʔc	*ʔc	ʰč
s	s	ʔs	ss	ʔθ	*ʔs	ʰs, s
š	s	ʔs	šš	ʔš	*ʔš	š, s
s	st	ʔn	ss	ʔθ	*ʔθ	š
s	hy	ʔn	ss	ʔθ	*ʔl	š (n?)

Examples of the PCA—Natick correspondences are cited below:

*mp : p · *brain, lift up.*

*nt : t · *enter* (see also *put into*), *hear, look, name, throat.*
: ʰt . *sharp.*

*nk : k . *armpit, dust, kick, star, sweet.*
: ʰk . *south.*

*nc : č . *come* (*proceed*) *from, finger.*

*ns : s . *view.*

*nš : s . *eye, nail.*

*nθ : ? no examples.

*nl : n . *name.*

There are examples in which the nasal cluster remains in Natick, cf. brain, related. The conditioning factors for these exceptions are unknown.

*hp : ? cf. *jaw, play.*

*ht : t. *forehead, make.*
: ʰt. *ear.*

*hk : k. *go after, grandmother.*
: ʰk. *deer, freeze, hold, louse, pot, root, short, spit.*

*hc : ? no examples.

*hs : s. *grandmother* (see also *sister* [1]).

*hš : ʰš. *so many* (?).

*hθ : š. *father, walk by* (cf. *sprout* for a possible exception to this correspondence).
: ʰs. *so many* (?).

*hl : š. *breathe.*

*Xp : p. *thick.*

*θp (PCEA *xp) : p. *back.*
: ʰp. *arm, sleep with, snow* (*v*).
(PCEA *θp) : sp. *taste.*

*θk (PCEA *xk) : k. *jaw, meet, shoe, work.*
: ʰk. *break* (*2*), *land, leg, pot, soften, teach.*
(PCEA *θk) : sk. *back, bark, bone, forehead, rock, seed, snake, top, uncover, watch.*
: šk . *woodchuck.*

*θk (PCEA *θk) : ˢ/šk . *bear, find, liver, woman* (see also *bitch, widow*). Cf. *sister* (2) for *Hk (*H : ?) : ˢ/šk.

*cp : ? no examples.

*ck : ? no examples.

*šp : šp . *above.*

*sk : sk . *eye, new.*
: šk . *fire, kick, meet, scrape.*
: ˢ/šk . *burst.*
: ʰk . *nail.*

*st : ? no examples.

*sk : sk . *lick.*
: ˢ/šk . *red.*

*ʔt , *ʔc : ʰt , ʰč . *big (2)* (see also *old person, sea, thumb, toe (big)*. Trumbull implies that this alternation is morphological in Natick; however, all the examples I found of *big (2)* occurring in combined forms point to phonological alternation.

*ʔt : ʰt. *fear, foam, tree, try.*

*ʔc : ʰc. *exhaust.*

*ʔs : s. *bitter.*
: ʰs. *fish, stone.*

*ʔš : s. *hair.*
: š. *eight.*
: ˢ/š. *big (1).*

*ʔθ : š. *between, bitch, fear, kill, three.*

*ʔl : š. *fly, net.*
: n(?) . *fly, place.*

NOTES AND REVIEWS

PEOPLE AND PLACES. Margaret Mead. Illustrated by W. T. Mars and Jan Fairservis; photographs. Cleveland and New York: The World Publishing Co., 1959.

C. F. VOEGELIN

This book is in a series which is explicitly written for children (A Rainbow Book), and the author's 'first book for young people' (p. 319). However, in the instance of Margaret Mead's effort, a book is produced which is equally useful both for children and for adults who are no more sophisticated than children in how and why anthropologists work in cultures other than their own. The first part (Man Learns About Man) and the last part (Man Asks About Man) generalize so successfully about anthropological attitudes that the book as a whole may well be the one book to cite in answer to the frequent question, what is the one book to read about anthropology? (The famous text books in general anthropology, from Tylor on, are introductory—serve as prerequisites for further and narrower interests—rather than as unique orientations wherewith to start and stop.) But, as would be expected, Margaret Mead has seen to it that no sensitive reader will stop reading in anthropology after reading the middle third of her book (How Some Peoples Live): The Eskimo (27 pp.), The Indians of the Plains (27 pp.), The Ashanti of West Africa (16 pp.), The Balinese (16 pp.), The Minoans of Crete (18 pp.).

The book ends with a page (319) ABOUT THE AUTHOR, who 'has made nine field trips, and learned seven different primitive languages.' Many anthropologists have made more than nine field trips; I know of no other who would admit to having learned seven different languages in the course of field work, though I do know of two who seem capable of such performance—Paul Garvin and Kenneth Hale. The talent is rare; those who have it should be able to teach the rest of us, for talent is at bottom successful technique. I quote that part of Margaret Mead's teaching relevant to this point in extenso (from the generalizations on anthropological attitudes, part I, in the third chapter, The Anthropologist at Work, but just pp. 84–86 and 88–92):

". . . a man watching another man can understand something about how he feels, and if he learns the other man's language, he can ask him questions and listen to his answers. So the study of human beings in many parts of the world began with men and women, trained to be anthropologists, who asked questions. At first they tried talking and listening through interpreters, people who know something about both languages, but this did not work out very well. The interpreters themselves often did not understand the questions and the answers, and so everyone was likely to get mixed up. To understand what really was meant, anthropologists had to learn the language the people spoke.

"There are several problems about learning the language of a primitive tribe which you might not think of right away. The first thing an anthropologist has to expect is that the new language will be quite different from European languages. That is, the differences between the new language and English will be greater than those between, say, English and French and Latin, or English and German and Norwegian, or English and Russian, because all these languages belong to the same family. When we are just beginning to learn the language of an unknown primitive people, we do not know whether it will be the kind in which there are only short words which always stand by themselves. We do not know whether the language will make the kinds of differences ours does when we use *she* to talk about a girl, and *he* to talk about a boy and *it* to talk about a stick. In languages like ours we can say: "*He* hit *her* with *it*." But in some other languages the same sentence would be "*Third person* hit *third person* with *third person*." Still other languages make more differences than we do, as they have special pronouns not only for males and females and things but also for long things and round things and hollow things. We do not know whether the word for two of something will be different from the word for one of that thing, and if it is, whether the difference will be like that between *cat* and *cats*, or *mouse* and *mice*,

or *ox* and *oxen*. We do not know whether the same words can be used for men and sheep and eggs or whether different words have to be used, such as *crowd of men*, a *flock of sheep*, a *clutch of eggs*, a *pride of lions*, a *swarm of bees*, or a *covey of quail*. We do not even know whether what we call "parts of speech" such as nouns and pronouns, verbs, adjectives and adverbs—will be anything like those in English.

[Citation of an Eskimo song that goes with a game cited by Franz Boas = ethnolinguistics]: "From this song you can see that Eskimo is a language in which one bit after another is added on to change the meaning of a word.

"When an anthropologist arrives among the primitive people he wants to study, he must find someone who can interpret at least a little, so that he can ask what people call things and how they ask questions and ask other people to do things. Probably no one there will have any ideas about what a word or a sentence is; nor will anyone know that their language has a grammar. They know how to speak it and how to correct children when they make mistakes. So the anthropologist must behave partly like a child who is being taught to speak and partly like the first person who made a grammar; and he must patiently collect a great many sentences: "I see a man, I see a woman, I see a dog; I see it (a woman), I see it (a man), I see it (a dog); on and on, until he can both speak the language and write down its grammar.

". . . The anthropologist understands a great many things about the savages which they do not understand themselves. But at the same time he can never feel exactly as they do, of course, and even though he speaks the language quite well, it is not his own language, his mother tongue in which he was sung to sleep when he was a baby . . .

". . . Yet this is the way the first anthropologists had to work . . . moving pictures and tape recorders had not been invented . . . His success in recording the language depended on whether his ear was sensitive enough to pick up sounds which did not exist in his language . . .

". . . And now we also have the moving picture camera and the tape recorder. Of course, it is still important for the anthropologist to use the fact that he is himself a human being like the people he is studying. The tape recorder cannot learn the language for him; the moving picture camera cannot make friends with anyone. But now for the first time he can bring back for other scientists to see and hear—and for himself to look at and listen to again and again—something that happened only once on a faraway island. As the microscope made it possible to see things too small for the naked eye and the telescope things too far away for the naked eye, film and tape make it possible to catch and keep more sights and sounds than the brain can take in and sort at one time

"So, today films . . . and tape recordings . . . have become the anthropologists' substitutes for experiments. In experiments, the scientist can do the same thing over and over—. . . In modern anthropology we still let history make the experiments for us, but we record them in such a way that they can be studied over and over. As our ideas grow and change, we can test the new ones out against the old films and the old tape recordings long after the people who were filmed and recorded have deserted their old ways and have come into the modern world."

The above strikes me as the simplest and briefest statement of the well known Mead position on the value of learning the language of the culture one is investigating. I agree with this in general; and at some points, feel competent to add to its definition; but at other points fail to understand what is meant.

(1) Interpreters—bilinguals who know the language of their own people better than the language of the anthropologist—translate the responses of monolingual informants to the questions put by the anthropologist. What subsequently appears in the scientific report, then, is a version of the monolingual informant's response, mediated by the translation of the interpreter. Such translation may be expected to include distortion because it is what I would call a 'reverse translation'—reversing the direction of coding. Rather than coding with benefit of an inventory which one has interiorized, the interpreter translates into the anthropologist's language, without the benefit of the inventory and categories which he learned while becoming a Kwoma (or whatever his culture may be). This particular point can be less briefly but better stated in terms of 'selection' and 'the original word game'.[1]

(2) The anthropologist who learns to

[1] Roger Brown, Words and Things (pp. xvi, 398. Glencoe, Illinois: The Free Press, 1958), reviewed by Harry Hoijer in Language 35.496–503 (1959); C. F. and F. M. Voegelin, Selection in Hopi Ethics, Linguistics, and Translation (in press, Anthropological Linguistics, 1960).

speak a very different language does precisely the same thing when he addresses a monolingual informant in the latter's language—he makes a 'reverse translation' and thereby introduces much the same kind of distortions that the interpreter makes when he translates from his native language into the anthropologist's. In learning the language of the people one is investigating, 'reverse translation' occurs when one talks to the people, but a less distorted kind of translation is possible from the utterances of the people talking to each other or to the anthropologist.

(3) The latter kind of unidirectional translation, which goes inwardly, utilizing interiorized inventories and categories, is possible today without learning to speak the native's language—with the use of a tape recorder and ancillary informants. That is, the original informant's fast responses are taped, and subsequently played back little by little, with the original informant's utterances repeatedly slowly by the ancillary informant; and so transcribed and translated by the anthropologist, guided by more than one informant.

(4) The use of the tape recorder for recording the language of the anthropologist who is learning the native's language may give us some cues as to why or how some succeed at this enormously difficult task, while others (most others) do not.

INDIANA UNIVERSITY

ON CONSONANTAL CORRESPONDENCES IN THREE ARAWAKAN LANGUAGES

DOUGLAS TAYLOR

Examination of such data on Arawakan languages as are accessible to me suggests that Island Carib, Lokono and Goajiro belong to the same subgroup, which probably should also include Paraujano and long extinct Taino.[1] Phonetically, Island Carib of the 17th-century Dominican dialect (abbreviated as IC) appears to have been much closer to Guianese Lokono of circa 1800 (L) than is modern Goajiro (G); but these last two languages have in common a considerably greater proportion (something in the order of 57% as against 46% for IC : L) of basic lexicon. Innovation,[2] contraction and assimilation[3] seem to have been far-reaching in the recent history of all three languages; and for this reason, forms cited below will be taken, in the main, from the earliest records to which I can refer.[4]

The several vowel systems are little understood (particularly so in the case of Lokono, for which *to see or look* is shown by different authors as: adika, adeka, adoka, aduka);

[1] See IJAL 24.153ff. So far as the present note is concerned, it matters little whether such a subgroup is or is not a reality.

[2] For some striking instances of phonological innovation, see my note On the history of Island-Carib consonantism in IJAL 26, no. 2, 1960.

[3] E.g.: IC binálepule > CAIC bináf(u)i *morning*, L adita > aita *know*, aditi > aiti *son*, biabite > bibiti > biti *four*, biama > bian *two*, ühükere > ekura *root*. Or compare L eke *eat* and eke *container*, *clothes* with their IC cognate equivalents, áika *eat* and ákai *container*; L kimisa *clothes* with its IC equivalent, kamísa, and with its Sp. model.

[4] For early IC, I have relied on Raymond Breton's 17th-century Dictionnaire caraibe-français and français-caraibe in Platzmann's facsimile editions (Leipzig 1892 and 1900), and the same author's Grammaire in the new edition by Adam and Leclerc (vol. 3 of the BLA, Paris 1878). For L, I have consulted the anonymous Arawakisch-Deutsches Wörterbuch and the Grammatik der Arawakischen Sprache in vol. 8 of the BLA (which also contains short vocabularies of Baniva, Bare, Caouiri and Yavitero; Paris 1882), The Arawak Language of Guiana by C. H. de Goeje (Amsterdam 1928), which includes the most elaborate comparative vocabulary of Arawakan that we have, and two articles by N. P. Hickerson in IJAL 19.181–90 and 20.295–301. For G, I must depend on N. H. Holmer's description of modern Goajiro in IJAL 15.45–57, 110–21, 145–58 and 232–35, having been unable to obtain the earlier work of R. Celedon (in vol. 5 of the BLA). Citations from other languages have come from de Goeje's comparative vocabulary (a supplement to which is to be found in his long article, Nouvel examen des langues des Antilles, in JSAP 31.1–120) and from various other sources too scattered to be enumerated here.

and only consonantal correspondences can be considered here, although such limitation may be held by some to be invalidating. Hickerson finds five vowel phonemes, i e a o u, for the modern dialect of Lokono; but says that [i] and [e] on the one hand, [u] and [o] on the other, vary freely in most forms, though they are distinctive in some; and that there are instances of morphophonematic alternation between 'stable' /i/ and /u/.[5] Under these circumstances, it might have been wiser to forbear; and had there been any prospect of new and better descriptions of languages belonging to this stock being published in the near future, I should have done so. As it is, I shall refrain from any attempt at reconstructing the phonology of a parent tongue common to this subgroup (which, after all, may never have existed), and confine myself to seeking, and trying to explain, the various correspondences among—as well as the developments within—the languages concerned.

One cannot expect to recover with exactitude the phonological inventories of the older dialects; but for the purposes of this note, Island Carib, Lokono and Goajiro may be taken to have the following phonemes and prosodic features: IC. m b p n nh t s k h l r i u e o (= [ë]) a V̨ (nasalization of vowels) and V ('free' stress), L. m b p n d t s k h l r i ü(?) u e o a, G. m p w n t c s ŝ y k ʔ h l r i i u e o a V̨, together with consonantal and vocalic length, shown here by double letters (CC and VV). The modern Central American dialect of Island Carib (indicated in citations by CAIC) differs phonologically from that of Dominica (IC) in lacking the nh-phoneme (or sequence?), and in having four additional phonemes: d c g. Forms cited from other languages will be spelt as in the sources from which they are taken.

A list of likely cognates follows; Island Carib (Dominican dialect), Lokono and Goajiro being shown in that order, and preceded by the English translation.

[5] Personal communication.

after -árisi : -adiki : -aciki, *arm or wing* aróna : adena : atina, *ashes* balísi : balisi : pari, *bark or skin* -úra : -üda : -tah, *bat* (*quadruped*) búliri : buhiri : posici, *bird* ulíbi̧ų : kudibiu : ucii, *blood* íta-ho, -íta : ütü, -ite-na : -iŝa, *bone* ábu-hu : abuna-hü : himpu, *capsicum* áti : hati : haŝi, *chop* (*or fell*) ásuka : asuka : aŝota-, *claw or nail* -úbara : -ubada : -upatau, *drink* áta : ata : asa, *ear* arík̨ai : adike : ace?e, *eat* áika : eke : ika-, *eye* áku : akusi : a?u, *flesh* ékriko (CAIC ógorogo) : isiruku : i?iruku (so Holmer; but de Goeje cites ŝiruku), *fly* (*vb*) amą́hara (CAIC ahámara) : amurda : awatti-, *foot* ukúti : ukuti : -?uih, *four* bią́buri : biabite > bibiti > biti : pienci, *from* (a sfx) -seę : -se : -hee, *hand* úkabu : ukabu : ahappu, *he* likía : likia : nia, *head* ísik, isóko : isi-hi : -ikii, ekiwi, *here* iáka : iaha : yaa, *honey* mą́ba : maba : mapa, *husband* iráiti : ireti : eeci, *juice* -íra : -era : -ira, *leaf*, *liver* -ubána : -pana, *maize* márisi : marisi : mariki > maikki, *moon* káti : kati : kaŝih, *nose* ísiri : isiri-hi : -ici, *one* ábana, ábą : aba- : wanee, *path* mábu : uaburuku : wopuh, *say or speak* ar̨aka : adiaka : aaŝaha, aiŝaha, *sea* balána, baláua : bara : palaa, *she or it* tukúia : tureha : sia (western dialect), hia (eastern dialect), *sleep* aruka (CAIC arúmuga) : adunka : atuka, *stone* i-síba-ni (meaning restricted to *stone used as anchor*) : siba : ihpah, *tail* íli : ihi : -sih, *that-masc.* (*one*) líra : lira : cira, *that-fem.* (*one*) túra : tura : tira, *there* iára : iara : caa, *they* nhakía (CAIC hagía) : nakia : naya, *this-masc.* (*one*) líka : lihi : cii, *this-fem.* (*one*) túka : tuhu : tiĥ, *thou* bukúia : bokia : pia, *tobacco* iúli : iuli : yiri > yiih, *tongue* inínie (CAIC i̧ȩ́ie-i) : -ie, ü-ie-hi : iyee, *tooth* ári : ari : ari > aai, *two* bíama : biama > bian : piama, *we* uakía : uakia : waya, *with* -úmą : -uma : -umaa, *woman* i̧aru > hi̧aru (CAIC hi̧áru) : hiaru : hiieer, *ye* hukúia : hukia : hia.

This list may be supplemented by others showing some likely cognates found in only two of these languages; so:

IC : L. *agouti* píkuli (CAIC fíguri) :

pukuleru > hokorero, *ask* (*inquire or require of*) álakota : ahadakuta > adakota, *boil* (*tumour*) sibúli : subuli, *chew* asákura : asakuda, *cut* íbika : ibika, *dig* asíka : atika, *face* isíbu : isibu, *fat* (*grease*) -akéle, ákolię : ikihi, *flame* eléleni : eheludun, *give* ísika : isika, asika, *grass* kálau : karau, *hammock* ékora (CAIC -ógora) : ukura-hu, *hear* akąba : akanaba, *kill* ápara : apara, *leg* -úrna : üdana, *name* íri : iri-hi, *pot* rúara : duada, *root* ílakra : ühükere > ekura, *scrap*(*s*) -ípe : -ipe, *see* aríka : adika, *slap* apátara : apatada, *slice* (*vb*) elésura : eheüsa, *stink* ínsi- > hį(n)si- : hisi-, *worm* (*maggot*) -íse : ise-hi.

IC : G. *flower* íleue : -sii, **siwi*, *heart* ánisi : ani > aį, *lip* iúmaru : iimata (in IC, a compound: *mouth-border*), *night* ariábu : ali > aih (cf. IC arábu *forest*, containing *ar-tree*, L uniabu *water*, containing uini *river*, *rain*, etc.), *sun* kási : kaih, *sweep* (*with broom*) abuíta : epitta, *what* káta : kasa.

L : G. *all* -makua : puŝua, *ax* baru : pore, *come* anda : ent-, *die* ahuda : au(k)taa (the k is present in Holmer's record but not in that of Celedon), *father* -ti : -ŝih, *fire* hikihi : sikih (for Trinidadian L, Dudley recorded hiket in 1594–95), *fish* hime : hime, *flow* amalada : aparala, *forest* (L) kunuku : *tree* (G) unu?u, *hair* ubara : -wala, walaŝ, *horn* ukoa : -uuwa, *I* dakia : taya (as against IC nukúia), *know* adita > aita : atiha (cf. IC aríta *think*), *mother* -iu : -eyu, *neck* unuru : anuulu, *potato* (*sweet*) haliti : haiŝi, *sand* mutuku : muaku, *seed* isi-hi : -si, *three* kabuhin > kabun : apuni, *walk or go* akuna : a?una, *water* (*substance*) uniabu, (*occurrence in nature*) uini, uin : wiį, *wife* iretu : eeriį.

Short as are these lists (54 items in three languages and 53 in two), they could not, I think, be usefully extended from the data at my disposal;—by which I mean that such additions as I could make would only multiply instances of adequately attested correspondences without clarifying those that are obscure (as, for example, are the Goajiro reflexes of p, t and s in Island Carib and Lokono).

In the paragraphs that follow it has, I hope, been possible to explain some apparent irregularities (e.g., m : u : w in the words for *path*) by different developments internal to the languages concerned; to accept others provisionally, on the strength of the evidence, while awaiting an explanation (e.g. 1 : h : s in the words for *bat*, *tail*, etc.); and to reject a few (e.g., apparent s : k in the IC and L words for *after* and in the L and G words for *head*) as incompatible with other and weightier evidence.

IC m : L m : G m (see under *honey*, *maize*, *two*, *with* and, for IC : G alone, *lip*, for L : G, *fish*, *sand*. Apparently irregular in this respect are the forms meaning *bone*, *fly*, *path* and, for L : G alone, *all*, *flow*; but an explanation can be given for at least two of these items.

G himpu *bone*, when compared with such other Arawakan equivalents as Mehinacu and Waura inapü, is seen to result from syncope of the vowel, a, and assimilation of the nasal to the labial articulation of the following stop. This nasal—or the morpheme containing it—has been lost in the IC and L equivalents, unless the latter should be a case of inversion (L abuna- < *nabu).

Also the initial phonemes of the words for *path* almost certainly result from assimilation (IC mábu < *nábu, L uabu- < *iabu- and G wopu- < *yopu-), preceded, in the case of the last two languages, by palatalization and loss of *n (L *iabu- and G *yopu- from earlier *inabu or *inapu); cf. Bare dinabu, Maipure anepu and Piapoco ayapu, all meaning *path*. Similarly, compare IC míbi with L nibi > mibi, both meaning *bush-rope*, *liana*, and IC mábi *sweet-potato* from earlier nábi ("a este *age* llaman los de Caribi *nabi*" Chanca and Columbus, Guadeloupe), itself borrowed from Karina napi, which has the same meaning. For palatalization and loss of *n after immediately preceding i, see under n : n : n below.

As regards the apparent correspondences of IC and L m to G w in the words for *fly* and of L m to G p in those for *all* and *flow*

I can only suggest that this may be the outcome of dialect mixture within Goajiro, which is said to be common. Thus, Holmer cites from the speech of different informants: miicka = piicka, waima = maima.

IC b: L b : G p (see under *ashes*, *bat*, *bone*, *claw*, *four*, *hand*, *honey*, *leaf*, *path*, *sea*, *stone*, *thou*, *two* and, for IC : L alone, *boil*, *cut*, *face*, *hear*, for IC : G, *sweep*, for L : G, *ax*, *three*). Seemingly irregular in this respect are:

G ucii (IC ulíbịų, L kudibiu) *bird* and -wala (L -ubara) *hair*, in which the labial stop appears to have been lost without compensation, and G wanee (IC ábana, L aba-) *one*. So far as the last two instances are concerned, G p may possibly have been replaced by w through non-distinctive voicing of the former phoneme, and its subsequent confusion with the w that replaces Sp. b in some loanwords like wooliiwa *bolivar*, kaawa *cabo* and tawaak *tabaco*.

G pienci (IC biąburi, L biabiti) *four* (all containing, as first element, morphs meaning *two*, q.v.). Here we seem to have two assimilations due to vowel syncope: medial *p of expected *piem(V)p(V)ŝi being replaced by t, and its *m by n, as in pientŝi or, since c is a phoneme, pienci.[6]

IC p (> CAIC f) : L p (> f) : G h(?). So far as IC : L is concerned, see under *agouti*, *kill*, *scrap(s)* and *slap*. But in neither language is this phoneme found in a large number of morphemes; and unless G alihuna *stranger* should be cognate with IC kalípuna and L kalipina, both meaning *Carib*, no likely correspondence for it can be shown for Goajiro. However, cf. G er(i) (L peru) *dog*, which was probably borrowed, like its L equivalent, from Spanish (rather than being cognate with IC áųli, Piapoco auri and other similar words for *dog*, which were inherited).

IC n : L n : G n (see under *arm*, *leaf* and, for IC : L alone, *leg*, for L : G, *come*, *forest*, *neck*, *they*, *three*, *walk*). Other instances are not hard to find; so, for IC : L, áunaba : aonaba *answer*, anéka : anika *take or carry*, anúata : anoa *choose*, anáko : anakü *high sea*, or *middle* (*of a place*), -áni : -ani *own* (*property*), néhene : nahali- *paddle* (*n.*). (I take it that final -ne and -li of the last example once belonged to separate and non-cognate morphemes; cf. Taino nahe, Bare neheu, both meaning *paddle*.)

As in the cases of G himpu *bone* and pienci *four* (see above), vowel syncope brings about assimilation of a nasal, or its reduction to vowel nasalization (see examples under *sleep* and IC : L *hear*).

In all three languages, n tends to be palatalized, reduced to vowel nasalization, or lost where i precedes (for examples, see under *tongue*, *woman* and IC : L *stink*; and compare with inene *tongue* and inaru *woman* of Carutana and other languages). In Lokono, all apical consonants may be palatalized when followed by i (as in -ani, hati, etc.), without thereby losing their phonological identity; while in Goajiro, some at least undergo regular phonological change in this position, as in wiį < *wini (L uni-) *water*, aih < ali *night*, yiih < yiri (L iuli) *tobacco*, -ʔuih < *-ʔuŝi (L ukuti) *foot*.

IC nh (>CAIC h) : L n : G n, as in: nhakía (CAIC hagía) : nakia : naya *they*.

IC r : L d : G t, c (see under *after*, *arm*, *bark*, *claw*, *ear*, *sleep* and, for IC : L alone, *chew*, *fly*, *leg*, *pot*, *see*, *slap*, for IC : G, *lip*, for L : G, *come*, *die*, *I*, *know*). On the basis of comparison with other languages of the stock, it seems justified to group these two sets together, and to conclude that the G opposition of c to t results from internal and fairly recent developments.

The forms meaning *bird* and *say* should probably be included here, although they show, respectively, l : d : c and r : d : ŝ. Interchange between l and r will be discussed below. Transposition or absorption of the i in the forms meaning *say* would account for G c rather than t; while loss of occlusion, resulting in ŝ, may well be due to the consequent opening of the syllable.

[6] As will be seen below, G c could here come from an earlier liquid (l or r); but in that case one must conclude that the second labial stop was quite lost.

This correspondence should be compared with that of l : h : s, for which see below.

IC t : L t : G t, as in túra : tura : tira *that-fem.* (*one*), túka : tuhu : tiï *this-fem.* (*one*) and (IC : G) abuíta : epitta *to sweep*. No other instance of this correspondence was found in Goajiro.

IC t : L t : G s, ŝ (see under *blood, capsicum, drink, moon, she* and, for IC : L alone, *ask*, for IC : G, *what*, for L : G, *father*, (*sweet*) *potato*. The same remarks as were made above with regard to G t and c apply to the grouping together of G s and ŝ.

Anomalous and unexplained are: apparent IC s : L t in asíka : atika *dig*, for which I can find no likely cognate in Goajiro, and IC r : L t in biąburi : biabiti *four*, whose Goajiro equivalent, pienci, has already been discussed under b : b : p above.

What looks like another instance of L (and in this case also IC) t : G c may be seen in IC iráiti : L ireti : G eeci *husband*; though here again I think that G c might be resolved into t + ŝ; since l and r of this language are regularly replaced by its t before another consonant (so, hiieer *woman* + -kal > hiieetkal, and er(i) *dog* + -ka > etka); and eer- (by metathesis from ire-) + -ŝi would give eetŝi, or eeci. (Alternatively, G eeci *husband* may be unrelated to its IC and L equivalents; and another possible ancestry for this form will be suggested under r : r : c below.)

Expected s or ŝ of Goajiro may be lost or reduced to h in some forms; cf. G muaku (L mutuku) *sand* and G atiha (L adita) *know*. But then it should be borne in mind that, dialectally, eastern G h and western G s or ŝ replace one another in some words at least; so: hia or sia *she, it*, hV- or sV- *she or her, it or its*.

IC s : L s : G s, ŝ. See under *chop* and, for L : G alone, *seed*; though in neither case is the Goajiro form indubitably cognate with the others. On the other hand, IC s : L s is well attested (as in the forms meaning *ashes, boil, chew, face, from, give, head, maize, nose, stink, stone*).

IC s : L s : G h. See under *from, stone* and, for IC : G alone, *sun*. The first, a suffix, occurs as in IC íniųseę *from on high* (ínių *high*), L hururuse (*out*) *from the ground* (hururu *ground*), G wattahee *from afar* (wattaa- *far*); and contrasts with the nominal dependent suffix (IC -te : L -te : G -see), occurring as in IC niúlite : L daiulite : G tayii(h)see *my tobacco*. The other two examples present rather a special problem (and that whether the consonant cluster of G ihpah *stone* results from vowel syncope or from metathesis); for Holmer says that syllable-final -(i)h of Goajiro often represents "original" -li, -ri or -ci, as in yiih < yiri *tobacco*. It would seem, however, that it might also represent earlier -si or -ŝi, as in the Goajiro forms meaning *foot, stone, sun*; (cf. also the L and G words for *fire*).

Such apparent correspondences to L s as G c (see under *nose*), G k (see under *head, maize*), IC k and G ʔ (see under *flesh*) must be rejected. As I have tried to show below, c of the Goajiro word for *nose* most probably developed out of an earlier liquid (l or r), and corresponds to r of the IC and L equivalents. The L and G forms meaning *head* obviously cannot be cognate if, as I believe to be the case, L isi-hi *head, tip, top* is identical (rather than homonymous) with L isi-hi *seed*, and if G -si *seed* is a cognate of this form. As I have previously suggested[7] the IC forms meaning *head* may have contained a morph belonging to IC óko *core, kernel*, etc., and their G equivalents a cognate thereof, together with a cognate of L iui *fruit*; and a similar analysis seems at least possible in the case of the words meaning *flesh* and *maize*. And if this should be so, then all these forms, except L isi- *head*, etc., were compounds or at least complex; the IC and L words for *maize* being fully cognate but only partly so with their G equivalent; the IC and G words for *flesh* being fully cognate but only partly so with their L equivalent; while L isi- *head* is cognate with the first part of its IC equivalent, whose second

[7] See IJAL 24.78.

member is cognate with the first part of G ekiwi *head*; the non-cognate morphs containing respectively s and k having probably been synonymous or nearly so This interpretation finds some support in the fact G apparently has or had two words meaning *flesh*: i?iruku and ŝiruku.

IC k : L k : G k, ?. See under *ear*, *eat*, *eye*, *flesh*, *foot*, *moon*, *sleep* and, for IC : L alone, *agouti*, *ask*, *chew*, *cut*, *dig*, for IC : G, *sun*, *what*, for L : G, *fire*, *forest*, *sand*, *walk*. There is no indication that the glottal catch of G developed out of a positional variant of the dorsal stop; thus, compare L -luku : G -lu?u *in* with L -ruku : G -ruku in the words meaning *flesh*, or the glottal catch in the G forms for *eye* and *walk* as against the dorsal stop in the G word for *sand*, which has the same environment. Nevertheless I do not believe, as does Holmer, that this fact necessarily indicates "the earlier existence of two k-sounds in primitive Arawak." As I have tried to show elsewhere, k of early IC split into aspirated ĸ and unaspirated g of the same island's more recent dialect, developing into h, voiced g and voiceless k (a rare phoneme) of CAIC; so that there is now no indication of why Breton's k (as in his "naríkiẽ *je regarde*" and "natíkiẽ *je pêche*") should have given CAIC h in one instance (naríhią), but g in another (nadígią), or of why his "eúkê *noyau, & tout autre chose qui a une enveloppe*" should have split into CAIC -óho *extract, essence* and -ógo *nucleus, vulva*.[8]

IC k : L k : G h (see under *hand*, *say*).

IC k : L k : G Ø (zero). The glottal catch of Goajiro "is not found initially or before a consonant and is weakened in word-final position" (Holmer). This would explain the absence of an initial consonant in the G forms meaning *bird*, *three* and *tree* (cf. also IC ulíbịų *bird*; though CAIC guríbiua, the name of a species of bird, would seem to be its cognate). Elsewhere, loss may be due to palatalization, as in the G forms meaning *he*, *I*, *they*, *thou*, *we*, *ye* (whose IC and L equivalents have a morpheme -kia or -kuia in common), or, it would seem, to labialization, as in the G word for *horn*.

IC k : L h : G Ø (zero). See under *here*, *this-fem.* (*one*), *this-masc.* (*one*). An analogous reduction has affected these same forms' cognates in the other two languages; cf. recent IC iáha, túha, líha; CAIC iá, toa [tɔ:], lea [lɛ:]; modern L ia, li, to.

As in the case of the apparently 'irregular' correspondences to Lokono /s/ discussed above, I must reject the apparent correspondences of IC s to L k and G k, as in the forms meaning *after*, and of G ŝ or s to L k, as in those meaning *all* and in L aküleka : asiliha *extend towards, offer, give* (cf. IC asilera *arrive*); and conclude that the morphemes concerned are not cognate, though the rest of the forms into which they enter may well be so.

IC h : L h : G h (see under *woman*, *ye* and, for L : G alone, *fish* and *sweet potato*). Rather expectedly, this phoneme seems to be unstable in all three languages. Recent IC and CAIC both have h corresponding to L h as in: hába : haba (*kind of*) *basket*, haburi- : haburi- *ashamed*, hála : hala *seat* (*n.*), hínsi- or hịsi- : hisi- *stink*, hịaru (IC) and hịáru (CAIC) : hiaru *woman*; though all of these have vowel initial in Breton's transcription of early IC. This may not have been due to negligence on Breton's part; for he discusses the aspirate, and shows it both initially and intervocalically in some other forms where h is today organic (e.g.: hukúia > CAIC hugúia *ye*, hamą- > CAIC hamą- *fly*, whence amąhara > CAIC ahámara *to fly*); while his áti > CAIC áti (as against L hati and G haŝi) *capsicum* both lack it, as does CAIC ą (sa) . . . ? *what* (*or how*) *about* . . . ?—supposing this form to be cognate with G hą < hama, which has the same sense.

IC l : L h : G s (see under *bat*, *tail* and, for IC : L alone, *ask*, *fat*, *flame*, *root*, *slice*, for IC : G, *flower*, for L : G, *fire*. Though some of the instances cited may be suspect

[8] See IJAL 24.238, and especially IJAL 26, loc. cit. in footnote 2 above.

(thus l of the IC form for *ask* might correspond to either the h or the d of its L equivalent), this correspondence—whatever its ancestry[9]—seems adequately attested by the remainder. So, some other Arawakan words for *bat* are: pisiri (Yucuna), pitseri (Cata- politani), pitiri (Carutana & Siusi), ŝiepir

[9] Although there is some indication that this correspondence may be confined to positions before what was a high or a mid front vowel, its occurrence is not to be attributed (as is that of G ci corresponding to li or ri of IC and L; see below) to factors pertaining to the internal development of this or that language. Nor can it be ascribed to regular sound change operating in two of these languages on a phoneme that remained unaltered in the third. The PARENT phoneme no longer exists; but clues as to its nature, though slight, are not altogether lacking.

In the years 1594–95 on the island of Trinidad, the Englishman Robert Dudley recorded as ending in "th" and in "t" several Lokono words ("addoth", "dasereth", "sibath", "tauraroth", "callit", "ureit") which lack these endings in all subsequent records (whose transcriptions indicate forms we write: ada, dasere, siba, kaiuru, kali, iuri or iuli). And some fifty years later on the island of Dominica, the Frenchman Raymond Breton recorded with final "t" several Island Carib words ("arĺbelet", "bichet", "boúrrélet", "hébechet" or "ſbichet") whose modern descendants and (or) cognate equivalents in closely related languages lack any reflex of this ending; so: CAIC alĺbele and L idibale, Taino bixa (as against Machiguenga pušote, ačiotl of unrelated Nahuatl), Taino burén and L budali, CAIC hĺbise and Taino hibiz. There is nothing very remarkable about the weakening and loss of consonants in final or in another implosive position; but in view of Breton's description of an Island Carib speech sound as an apico-interdental lateral (*v. infra* in text for his own words), preceded, in the speech of a minority only, by an apical occlusion, Dudley's "th" of Trinidadian Lokono is of interest.

It suggests the earlier existence of a phoneme not unlike the tl of Nahuatl, which was eliminated from word-final position in at least two of the languages, but was merged in some other positions with IC l (through spirantization and loss of the first element), L h (through aspiration or spirantization of the first element and loss of the second) and G s (through assibilation of the first element and loss of the second).

It is tempting to seek another instance of this correspondence in the markers of the nominal absolute, L -hV (vocalic harmony) and G -ŝ, as in L abunahü *bone* vs dabuna *my bone*, G walaŝ *hair* vs tawala *my hair*,—especially as the Nahuatl suffix -λ(i) has the same function! But that is improbable; for IC has -ho ~ -hi ~ -hu (not *-lV with the same function as the L morphs; only G (but not ŝ) is found in other instances of this correspondence; the L and G absolutive suffixes do not both occur with cognate stems (thus: L ubana *hair*, G himpu *bone*).

On the other hand, comparison with cognates from outside this group (e.g., Baure huitere *bat* suggests that only where the 'parent' phoneme was between high or mid front vowels—or, when word-initial, was followed by such a vowel—did it yield IC l : L h : G s; and that in other environments it gave IC r : L d : G t, c. Thus Carutana for example, has the same apical stop in pitir (IC búliri : L buhiri : G posici) *bat* or itipi (IC ſli : L ihi : G -sih) *tail* as in tawana (IC -úrna : L -ūdana) *leg*, and Manao has the same apical stop in ghügaty (L hikihi : G sikih) *fire* as in ata (L ada) *tree*, -nyupâta (IC -úbara : L -ubada : G -upatau) *claw* or -teki (IC arĺkai : L adike : G ace?e) *ear*. If this should be correct, it would seem that the IC, L and G words for *bat* have lost a front vowel, or replaced it by a back vowel, and that IC l in the word for *ask* corresponds not to the h but to the d of its L cognate equivalent (as in the case of the words meaning *bird*). Moreover, though Palicur -ribu *tail* and tiketi *fire* can hardly both be cognate with their IC, L and G equivalents (cp. also Palicur -riui *fruit* with the IC and G forms meaning *flower*), the former language does seem to have undergone a change analogous to that which presumably produced our two positional sets (l : h : s and r : d : t, c); for its h regularly corresponds to L d (IC r, G r, c), as in aha *tree*, kuhipra *bird* and himaka *sleep*. But here great caution is needed, particularly in the case of Palicur t, which apparently stems from more than one phoneme of the parent language, and corresponds to L t (as in atiti : hati *capsicum*), L s (as in tipa : siba *stone*; cf. Carutana hipa and Manao ghüa *stone*), and L h (as in tiketi : hikihi *fire* or zero (as in pitana : biama *two*; cf. Piapoco putsaipa, Carutana ndzame, Ariti hinama and Taino yamoka, which have the same meaning). Also IC l may correspond to zero of Lokono, as in liki- or like- or leko- : L -iki-n- *only, alone*; with which compare Taino heketi *one*. In this case at least, the absence of an h in the Lokono form may be explained by the fact that it does not apparently, occur without a pronominal prefix, whereas with that of Island Carib, which is itself often enclitic, only a suffix (pronominal or other) is found.

(Ipurina); and for *tail*: ihibi (Bare), iŝipi (Ipurina), itsipi (Catapoliyani), itipi (Carutana & Siusi), -ribu (Palicur), enihu (Ariti, which lacks a labial stop); while from L ihi and IC íli *tail* are derived the verbs: L ihika, which the Herrnhuter author of the Arawakisch-Deutsches Wörterbuch modestly translates "*concumbere con muliere*", and IC aliákua, which has the same meaning. In IC ílakra : L ühükere *root*, IC l can correspond to nothing else but h of the Lokono (lost in modern ekura); and cf. Wapisiana luakare *firewood*, a probable cognate. No absolute form of G -sii *flower* is given by Holmer; but on the analogy of G -ikii, ekiwi *head*, and of IC uíra : L iuida : G iita *calabash*, one might expect it to be or have been *siwi, corresponding to IC íleue, Adzaneni eliwi and Bare hiwi, meaning *flower*, Palicur -riui and L iui *fruit*. L hikihi and G sikih (from earlier *sikisi?) *fire* have such other likely cognates as: tidze (Carutana, Siusi & Catapolitani), tiketi (Palicur), tikiahi (Saraveca), ilika (Ariti), all meaning *fire*, and Guana tikoti *tree* while IC uelékia *fire-drill*—with which compare Breton's IC "liki-likiti ou riki-rikiti, *branlé, ébranlé*"—and (or) íleme *blaze, kindling* are at least possible kin.

IC l : L l, r : G r, l (see under *ashes, sea, tobacco* and, for IC : L alone, *agouti, boil, flame, grass*, for L : G, *flow, hair, neck*). The modern dialects of all three languages are said to have two liquids, l and r; and this opposition was already present, so far as I can tell, in early IC (cf. íli *tail* but íri *name*, íla *its contents* but íra *its juice*). The Herrnhuter Grammatik der Arawakischen Sprache even claims (but does not write) a third liquid "zwischen *l* und *r*, den man sehr schwer nachsprechen kann"; and Breton may have meant to indicate something similar for IC by his double "ll" (employed inconsistently, and not distinguished here from simple l, since apparently non-distinctive), or by his "dl" in "dleuti" (and only one other morpheme), on which he comments: "la plus part disent *leuti*, mais cêt *l* se forme par le bout de la langue engagée entre les dents." Nevertheless, l and r are given as interchangeable in some forms (cf. that cited from IC at the end of the preceding paragraph, and L "úkkulu, úkkuru, *das Knie*", "wurráru, wullálu *Wolken*,[10] *Nebel*"); and minimal pairs showing this opposition are extremely few (for Lokono the best I can find are: "kálli *Cassabi*" vs "kári-hi *Krankheit, Schmerz*", "ullúku *in*" vs "úrruku *Nymphae*"; which are valid only if we suppose stress and gemination or length to have been non-distinctive, as appears to be true of the modern dialect).

On the other hand we find, as between earlier and later dialects both of Island Carib and of Lokono, besides l > l and r > r, l > r (perhaps especially, but neither only nor always, Breton's "ll" > r), but only very exceptionally r > l (IC aríbelet > CAIC alíbele *roasted meat* is my only example). So, for instance, IC iúli : L iuli *tobacco* and IC balána or baláua : L bara *sea* have become, respectively, recent IC and CAIC iúri : modern L iuri, recent IC and CAIC barána or baráua : modern L bara.

IC l : L l : G n. This correspondence—if it be one—was found only in the words meaning *he*, and in the prefixes meaning *he, his*. I can offer no explanation.

IC r : L r : G r (see under *flesh, juice, maize, that-fem., that-masc., tooth, wife, woman* and, for IC : L alone, *hammock, kill, name, root*, for L : G, *ax*).

The development of G mariki > maikki *maize* may perhaps be explained as the product of metathesis (mari- > *mair-), morphophonematic alternation (*mair- + -ki) *maitki) and assimilation (*maitki > maikki), in the same way as we have sought to explain the Goajiro word for *husband*.

G eerįį *wife* (cf. L iretu) may likewise result from ire- > eer- + *-ni > -įį; and if so, it is only partly cognate with its Lokono equivalent.

An l or an r may be lost in Goajiro, as in L haliti : G haiŝi *sweet-potato*, IC iára : L

[10] Cf. IC alíru *white cloud*, ualįų *black cloud*.

iara : G caa *there*, G ari > aai *tooth*, this latter with compensatory lengthening.

IC li, ri : L li, ri : G ci (see under *bat*, *nose*, *that-masc.*, *this-masc.*). The Goajiro reflex seen here is an OCCASIONAL correspondence due to various factors such as weakening, dissimilation and analogy. That it is not a regular sound shift appears from the occurrence of such a form as G pari (IC balísi, L balisi) *ashes*.

In such Goajiro forms as show it, an earlier liquid must have become palatalized and fallen together with y (cf., in closely related Paraujano, yi *nose* and eiy(e) ~ eitsch *man*; the latter apparently cognate with IC eiéri *man*). This y was then STRENGTHENED to [dž]—a sound which one of Holmer's informants, he tells us, employed "for initial y in yiih *tobacco* and in some similar words." However, in a language whose phonology includes a hushing palatal affricate, c, but no voiced—voiceless opposition, this sound could hardly escape identification with the affricate; and this has clearly taken place in the case of G caa (IC iára, L iara) *there*, which otherwise would have been indistinguishable from G yaa (IC iáka, L iaha) *here*. There is therefore no reason to suppose that the c of G cira (IC líra, L lira) *that-masc.* (*one*), G cii (IC líka, L lihi) *this-masc.* (*one*) and G posici (IC búliri, L buhiri) *bat* had a different origin.

Some doubt might remain as to whether the c of G -ici *nose* corresponds to the s or to the r of IC ísiri, L isiri-; but no other indication of the former correspondence has been found; Paraujano yi *nose* points to *iyi as the immediate parent of the Goajiro form; and, while Wapisiana iribe and Paumari iridi, both meaning *nose*, lack any reflex of s in the IC and L equivalents, no likely cognate that lacks a reflex of these same words' r has been found.

The Paraujano forms meaning *man* as cited above are sufficiently similar to G eeci *husband* to throw doubt on the latter's ancestry. In this form cognate with IC iráiti and L ireti *husband*, as we have suggested, or with IC eiéri (Palicur hiñeiri, Piapoco asieri) *man*? If with the former, one might expect the hypothesized metathesis (ire- > eer-) to remain stable in this morpheme, whose cognates are also found in L ire-nu *children* and IC (i)rá-ho *child*, *little* (*one*). But we have G ahapp-ira *finger* (*hand-child*) corresponding to IC -úkabu-rá-hị with the same meaning, where this transposition of the r has not occurred. It may therefore be that G eeci *husband* is cognate with the IC and other forms meaning *man*; and that G eerị *wife* contains a contraction (or one morph?) of G hiieer *woman* (cf. Palicur ero *woman*, Manao ero *wife*) together with some other morpheme (cf. IC and L -ani *belonging*, Maraua runo *woman*).

DOMINICA, BRITISH WEST INDIES

COMPOUNDS AND COMPARISON AGAIN

DOUGLAS TAYLOR

As I tried to show in a recent note under the same heading (IJAL 24.77ff.), comparison of not too distantly related Arawakan languages often reveals the presence in one of PETRIFIED former morphemes (both roots and affixes) that are still 'alive'—can be isolated, that is to say—in another.[1] It may then appear that some very similar equivalents in two languages, the one a compound and the other synchronically complex or even a simple form, are no more

[1] So, Lokono -itima *beard or moustache* appears to be unanalysable in that language (despite its uima *mouth of a river*); whereas its Island-Carib equivalent, -ítiuma, is readily resolved into -iuma *mouth* and iti- (cf. ítiburi *hair*, and biąburi *four* from bíama *two*) *hair*, *strand*, *piece*. Similarly, Island-Carib iráiti *husband* (despite iráho *child* of the same language) is not susceptible of analysis; whereas its Lokono equivalent, ireti, is readily resolved into ire- (as in irenu *children* with the nominal pluralizer -nu, ireiu or iretu *wife*) *child* and -ti *father* (as in dati *my father*; cf. daiu *my mother*).

than partly cognate.[2] In what follows, I call the reader's attention to a few further possible instances of both sorts, and try to amend what now appear to have been a couple of faulty interpretations on my part.

There can be no doubt that Lokono "idiussi *die Warze an der Brust*" and "jahússi *Kattun-Saamen*" are compounds of the same language's "ídiju *die Brust*" and "jáhu *kattun*" with "íssihi *der Kopf, das Äusserste, der Gipfel*: dássi, bússi, líssi, tíssi, wássi, hússi, nássi [*my, thy, his, her, our, your, their head*, etc.]; heisst auch: *Saamen*"; and that irrespective of whether the latter entry cites two homophonous morphemes, or only one as the author seems to imply.[3] And it might be justified to recognize its "iwi *die Frucht an allerlei Bäumen*", together with -si *seed*, in "íwissi *cuniculi, die Hoden*; auch wohl: *penis*; díwissi, líwissi oder dêsse, lêsse; íwissi lúkkudu *die Hoden*; íwitikittin *fruchtbar machen*; . . .". But then we also find: "issínihi *penis*: dássîn, bússîn, líssîn; tissina *id quod est in nymphis*; . . . ínissi *cuniculi*: dêsse, bêsse, lêsse; elonti ínissi" *little boy's testicles*; from which it appears probable that ínisi contains -si *seed* in common with its quasi-synonym íwisi; and at least possible that ínisi and isini- (-hi is a suffix marking the absolute noun) contain or contained the same two morphemes differently ordered. But if so, ini- and -ni must belong to a morpheme other than iwi *fruit*; though possibly cognate with Breton's Island-Carib "im" and "-im" or "-ijm", which he translated in different contexts by *fils, fruit, graine ou semence, oeuf*; and which, as hí, -í of the modern Central American dialect of that language, still has the same range of referents.

In the earlier note referred to above, I concluded that Lokono -wasini or wasina *heart* and its equivalents in Island Carib (ánisi), Goajiro (-aị), Guinao (inišini), Piro (-wahi), Saraveca (-vihine), Maipure (-nikini) and Manao (-niki or -neki) were all cognate; but I am bound to reconsider this position after finding the Herrnhuter author's Lokono "ánsi, ánissi, ánsiwa"—undefined, but apparently referring to the non-physical heart, as in the example of its use: "júra dánsi *ich verlange*",—more literally perhaps, *my heart* (dánsi) *urges or pulls* (júra). As in the case of ínisi *testicles* and isini- *penis*, so Lokono an(i)si(wa) *heart* = *own mind or soul or desire* may well contain or have contained the same morphemes, in reverse order, as -wasini or -wasina *heart* = *organ of the body* of the same language. But they are nevertheless different, apparently non-interchangeable constructions; and the whole problem is complicated by the existence of other possibly related Arawakan forms like Piapoco inasi and Ipurina isini, both meaning *flesh*.

In another note in this journal (25.195–96), I tentatively concluded that Island Carib, denominal attributive "kinchin *cher, aimé*" and privative "minchin", its antonym,

[2] For example, in view of such correspondences as are apparent in Goajiro iita, Island-Carib uíra and Lokono iuida, all meaning *calabash*, synchronically unanalysable Goajiro iimat(a) *lip*, whose parenthetic vowel is often omitted, might seem to be fully cognate with its Island-Carib equivalent, iúmaru, a compound of iuma- *mouth* with -aru *border or edge*. But this compound's second member is comparable with Lokono aruma *side, edge, coat*; and while Island-Carib /r/ corresponds to either /d/ or /r/ of Lokono, Goajiro /r/ corresponds to /r/ of Lokono, Goajiro /t/ to /d/ of Lokono. It is therefore clear that the Goajiro word for *lip* never contained a cognate of Island-Carib áru *border*, but, instead, a morph belonging to its own -tah (Island-Carib -úra, Lokono -üda) *skin*. Analogously Lokono, from its -leruku *mouth*, has formed -lerukuda *lip* (cf. also French Creole lapo buš *lip or lips*, but literally *skin* [*of the*] *mouth*).

[3] Citations of Lokono are from the Arawakisch-Deutsches Wörterbuch or from the Grammatik der Arawakischen Sprache, both anonymous and both contained in volume VIII of the Bibliothèque Linguistique Américaine; Paris, 1882. If Schultz and Schumann were the respective authors of these works, as is generally believed, they were most probably compiled during the last quarter of the eighteenth century.

whose stem became -isę or -isen in the recent Dominican dialect and -ísię in that of Central America, were cognate with Lokono "kanissin *lieben*" and "mansin *nicht lieben, hassen*", but had been reshaped under the influence of borrowed Karina ise *to want or to love*. I still think this influence most likely so far as the later developments are concerned; but in view of the occurrence in recent Dominican of such a form as: kanísiti núni *I like it*—more lit.: *it-is-agreeable to-me* (containing ánisi *heart, inclination or mind*), besides khíseti núni (containing -isę), with much the same meaning, I now find it more plausible to compare 17th-century Island-Carib -įsį or -insin (the status of nasalized vowels in the earlier recorded stages of this language is uncertain) with Guinao inišini *heart* and—semantic difference notwithstanding—Lokono ínisi *testicles* than with Lokono án(i)si and the forms meaning *love* and *hate* as cited above. It is of course not inconceivable that Island-Carib -insin and ánisi, Lokono ínisi and án(i)si (whence -anisin), should have resulted from the splitting of what once were the same morphemes in the same construction; but it seems clear that these forms were quite distinct in both languages at the time of our earliest records.

Moreover, Lokono -án(i)si and Island-Carib -ánisi—with which compare Goajiro -aį (<*-ani) *life, heart, soul*; a form that enters into expressions of love or desire—appear to contain -áni *own* (*property or concern*), a stem found in both the former languages,[4] together with -si *seed*. And if this should be so, it is easy to understand how such a compound might be employed in reference to the heart—either as a mental and emotional essence, as a bodily organ, or as both.

It is more difficult to reconcile the semantic aspects of the other pair of forms that I have assumed, perhaps wrongly, to be cognate. Breton's "kinchin *cher, aimé*" may, in all likelihood, safely be analysed as containing the attributive prefix kA-, a subordinative suffix -n, and a nominal stem -insi- which seems to mean *dearness*. But is the latter an etymologically simple form? If it is really cognate with Lokono ínisi *testicles* and (or) Guinao inišini *heart*, it most probably contains -si *seed*, etc., together with a morph that I can best compare to Breton's Island-Carib "im", hį and -į of the modern Central American dialect, and which might here be glossed *kin, kind, kindred*, or still more plainly, *progeny, profit*; (but cf. also Lokono uina, -ina and Island-Carib iuna, -ina, both meaning *base, foundation, stem*).

Lokono akusi *eye(s)* and akuke *eyelid(s)* appear to be compounds containing a cognate of Island-Carib áku and Goajiro a'u, both meaning *eye(s)*, together with, respectively, Lokono -si *seed* and eke *container, covering*. But can this cognate be identified with the only similar free form listed in the Herrnhuter Wörterbuch: "aku *in, an, auf, über*"?

It seems plausible—though perhaps unlikely—that Lokono and Island-Carib márisi *maize* should be a compound of those languages' mári *toothless* or, at least in the former language, *blunt (without edge or*

[4] So Lokono: "ani *es gehört, ist eigen*; tuhu dani *das gehört mir*; kia bani *je was dir*; tain (ta ani) *nach seiner Art*; . . . "; and Island-Carib: "ani *chose*", "kat ani lika? *à qui est-ce que cette chose appartient?*", "nani-kia *c'est à moy*", "nani mhem-kia *j'en ay affaire*", "ao kani lika *je prend celà*". In both languages it may be verbalized; so Lokono: "anin *tun, machen*"; and Island-Carib: "aní-bae lika *fais celà*". That it denotes, in Island Carib as in Lokono, *property* rather than *thing* may be gathered from Breton's own examples of its use, or from the following, which were recorded by me among the Black Carib of Central America: báni líra *that belongs to thee, that (is) thy property, that (is) thy affair*; or verbalized as in: haní-bai (iáragię) *take it (from there)*. With the basal suffix -(i)na of this dialect (v. IJAL 22.14), it means *cause*, as in: tánina iąuari *the cause of trouble, the root of evil* (iąuari *evil, wrong*); and the verb ánida means *to own, to possess 'en propre'*.

point), from ári *tooth, teeth*, together with Lokono -si *seed*. Another possible etymology for the first part of this name is Lokono mári (Island Carib míri) *nameless*, from íri *name* of both languages.

But although -si *seed*, etc., remains as a Lokono stem, there are indications that it has sometimes degenerated, even in that language, to a suffix, if not further. So, compare Lokono and Island-Carib balísi *ashes* with abalin- "*vorüber gehen, weggehen*" of the former language, balí-ba "*torche, nettoye*" of the latter, and with Piapoco bali, Goajiro pari, both meaning *ashes*; or Lokono and Island-Carib -ubúlisi *dust, spray, chaff* (according to context) with abuledin- "*wegwerfen*" and abuledunua "*eine Quelle*" of the former language, buli- *spring up out of* and ábulikani "*source*" of the latter.

In Island-Carib -isípisi-íra *gall*, containing sipi- *bitter* and -íra *juice*, the morph -si can have none but a purely grammatical function (nominalizer?), with which Lokono itself dispenses in its isipe *gall*, from sipe- *bitter*. Island-Carib (n)árisi and Lokono (d)adiki, both meaning *after* (*me*) or (*my*) *trace or footprints*, most probably contain the same, cognate morphemes as are seen in Island-Carib ári *navel* and Lokono adi *upon, over* (*in addition, more, apart*); but there are no grounds for supposing that -si and -ki of these words are cognate, though they may well have the same function or significance.

It may seem unreasonable to refuse to recognize a correspondence between sibilant and stop of Island Carib and Lokono (cf. also IC asíka and L atika, both meaning *dig*) in view of the fact that Island-Carib ánisi, -árisi, balísi, ébesi (*payment, recompense*) and -isípisi-íra of the Dominican dialect have become: anígi, -árigi, balígi, ebégi and -igífigirá in that of Central America. However, in the case of the Island-Carib dialects we have evidence of the splitting, in St. Vincent, of s into s and c (aníci, ébeci), and of the all but universal replacement by analogy, in Central America, of /ci/ by /gi/; whereas such apparent correspondences between Island Carib and Lokono are few and sporadic (see my note on the history of Island-Carib consonantism in IJAL 26. ff.).

Island-Carib aríkai and Lokono adike, both meaning *ear*(*s*)—that is, the external appendage so called; there being other words for the organ of hearing as such—most probably contain IC ari- and L adi-, as in the forms meaning *after* (see above), together with morphs belonging respectively to Island-Carib ákai and Lokono eke *container or covering*. Cf. Lokono akuke *eyelid*(*s*), mentioned above; de Laet's "saeckee" *mother*, recorded in 1598, modern -sake *uterus, eggshell*, containing -sa *child, egg*; and Island-Carib ulákai *belly*, containing, according to Breton, úle *food*.

Island-Carib (h)uríra and Lokono idiúra *milk* are transparent compounds containing, respectively, úri and ídiu *breast* together with -íra and -éra *juice*; these second members being found also in Island Carib áku-íra and Lokono ikíra, both meaning *tear*(*s*). In its two stresses and lack of vowel reduction, the latter Island-Carib form is more like a syntactic construction ("*eye juice*") than a compound; but in view of Lokono akuke *eyelids* (see above), we can hardly accept the first member of Lokono ikíra *tears* as an allomorph of aku-. This morph apparently occurs again in Lokono -ikíti, ikítihi *eyelash*(*es*) (cf. Island-Carib áku-íu *eyelashes*, containing íu "*poils, épines*"), for whose second member see footnote 1. It seems possible that this iki- may be related to "ikîn *regnen*" and (or) to "ikin *einzig*", in reference to the distinct nature of the several tear-drops and eyelashes (cf. also -ikísi *count, reckoning*, etc.).

Evidently a case of not seeing wood for trees is my own blindness, hitherto, to the full significance of Taino 'arcabuco' *forest* as a presumable cognate of its Island-Carib equivalent, arábu (Dominican) > árabu

(Central American). This significance is emphasized by a comparison of Island-Carib ariábu (Dominican) > áriabu (Central American) with Lokono orikahu, both meaning *night*. In ascribing to these Island-Carib words and to Lokono uniábu *water* the status of etymological compounds (IJAL 25.192), I was, it seems, quite right; but the historical complexity of the forms now appears greater than I then perceived.

Island-Carib arábu *forest* < ar(a)- (Lokono ada) *tree or trees* + -kabu *comprising* ± -ku *at*, with loss of the dorsal stop in the second member and omission of the third. This -kabu is no doubt cognate with the homophonous suffix of early Island Carib and with 'propensive' -habu of the modern Central American dialect (numbered 323 in IJAL 22.4 & 14); and most probably enters, as stem, into Island-Carib úkabu (Dominican) > úhabu (Central American) *hand(s)*, Lokono "Ûekkabbu *die Hand*, insonderheit das Vordertheil, die Finger". When we compare the latter part of this citation with the same author's "Abúttun *anfassen, greifen, halten*", and with Island-Carib ábuta *to pick up, to gather or take in*, it becomes clear that these verbs and the nouns meaning *hand* alike contain a form abu *with* (*bearing, holding, in possession of*), common to both languages, and itself having the grammatical behaviour of a noun (v. IJAL 22.36). The form kabu *taking in, comprising*, which must have entered also into Island-Carib ariábu *night* and Lokono uniábu *water*, is therefore itself complex, consisting of the attributive prefix kA- together with the nominal stem -abu; (cf. Goajiro aih < ali *night*, and Lokono úin, úini or úni *river, rain, water*). The final syllable of Taino 'arcabuco' is not found with the Island-Carib and Lokono forms cited here; but unless it should be an independent suffix, it may belong to the same morpheme as Taino '-aco' *eye* and be cognate with Lokono aku "*in, an, auf, über* ('*in über*' sonderlich vom Feuer und Wasser), . . ; akulukku *in* (wird vom Feuer und von den Augen gebraucht), . . "; cf. also Island-Carib áku *eye(s)*, Goajiro -aʔu *eye(s)* and -aʔu *on*, Lokono akusi *eye(s)*; or Lokono abu- *have*, but kabukü- "*kriegen, bekommen*".

Taino 'conuco' *cultivated plot or provision ground* is certainly cognate with Lokono "Kúnnuku *der Busch*" and Goajiro unuʔu *tree(s)*; but it seems very doubtful whether these words can be related to Lokono "akunnuku *zu gehen*".

DOMINICA, BRITISH WEST INDIES

RATE OF MORPHEMIC DECAY IN MECCAN ARABIC

ARNOLD C. SATTERTHWAIT

The validity of the theory of lexicostatistical dating depends on the hypothesis that the percentage "of the basic-root-morphemes of a language *which* will survive as cognates after 1000 years *is constant* for all languages at all times."[1] If two different datable time-periods in the development of the same language can be determined, the percentage of the basic-root-morphemes of the language which have survived per thousand years; i.e., the rate-constant of the formulas, can be found. Insofar as the rate-constants for the different languages are approximately the same, the theory is confirmed. Investigation of the validity of the theory involves the examination of as many languages as possible to determine the rate-constant of each language. The validity of the theory depends upon the degree of agreement between these individual rate-constants.

Meccan Arabic lends itself ideally to this type of investigation to which originally 13 languages were subjected.[2] This dialect furnishes a datable body of material, the Qoran, which was revealed to mankind

[1] Robert B. Lees, The Basis of Glottochronology, Lg. 29.119. The italicized words were inserted for this paper.

[2] R. B. Lees, op. cit., 118–119.

by an identifiable Meccan, the prophet Mohammed, between 610 A.D. and his death in 632. The text was standardized by a second Meccan, 'Uthman ibn 'Affan, between 645 and 650. It is generally agreed that the consonantal text of the Qoran has undergone little change since that date.

In the following note the rate-constant for Meccan Arabic is obtained by comparing the Arabic of the Qoran with the Arabic spoken by contemporary Meccans. For purposes of this investigation I have assumed that the Arabic of the Qoran represents the particular Meccan dialect of Arabic in 632 A.D. from which the dialect spoken by Meccans today is derived. The rate-constant obtained by this comparison will be compared with the mean rate-constant derived by R. B. Lees, .8048 ± .0176/mill.[3] and the more recent mean rate-constant established by Morris Swadesh at about 6% above Lees' through further refinement of the method.[4]

In order to estimate the validity of a word-list it is necessary to know how it was formed. I established the Qoranic word-list as follows. (1) I collected all Arabic words which I recognize as being equivalent to Swadesh's revised list of 100 semantic items.[5] (2) Checked this list against a Qoranic concordance.[6] (3) Generally selected the most frequently occurring item as the Qoranic equivalent of Swadesh's semantic term. Only occasionally did I investigate the exact meaning of a word by reference to the Qoran. In several instances this investigation led to the rejection of an item which occurred more frequently.

This method of obtaining a word-list can be expected to skew the results a bit. Words with which I am not familiar have, of course, not been considered. Such words might possibly occur more frequently in the Qoran than those known to me. The result of omitting these words from consideration would be a tendency toward a higher retention rate than should actually be the case.

The following list includes all the items investigated with the results tabulated. The first word in each instance is the one selected as the Qoranic equivalent. The number in parentheses following each item indicates the number of times it appears in the Qoran. For high-frequency items I have indicated only the number of columns in the concordance, for example (4C), needed to list all occurrences. (N) indicates the item occurs but is not listed. Items followed by no number were checked but do not occur.

1. *all* kull (47) ʔajmaʿ (22) kaaffatan (5) jamiyʿ (3) qaaṭibah, saaʔir, 2. *ashes* ḤUṬAAM (3) rumaad (1) hamiyd (1) ḤADIYD *sharp* (2) *iron* (5) ḤADIYDIYYAH, sakan, dardaar, xaraabah, ṬFW, jmr, FḤM, 3. *bark* none, jiðʿ *trunk* (3) ǧiṭaaʔ *veil* (2) qišr, liyf, laḤaaʔ jilaaqah, 4. *belly* baṭn (4) *womb* (3) *valley* (1) plus 7 additional occurrences not investigated, jawf (1), maʔnah, kars, kariš, maʿidah, 5. *big* kabiyr (42) ʿaðiym *mighty* (more frequent; rejected on meaning), samiyn (3) jasiym, jaliyl, ʿawd, DXM, 6. *bird* ṬAYR (18) ṭaaʔir (1) ʿusfuwr, 7. *bite* ʿaDD (2) ǧrD, garraD, srk, DRS, qazqaz, jazjaz, nhš, 8. *black* ʔaswad (2), 9. *blood* damm (7), 10. *bone* ʿaðm (2), 11. *breast* sadr (many times), nhd, bzz, zaaḤ, θdy, 12. *burn* ʔi Ḥṭaraq (1) šbb, 13. *claw* none, xlb, Dfr mnšr, Dlf, muxbaaṭ, 14. *cloud* siḤaab (9), ǧym, nwʔ, 15. *cold* baarid (2), 16. *come* ʔataa (15C), jaaʔ (10½C), ḤaDar (1C), 17. *die* maat (2¼C), tuwuffiy (1¼C) 18. *dog* kalb (6), 19. *drink* šarib (13), skr (5), 20. *dry* yaabis (2), jaaff, nšf, 21. *ear* ʔuðun (16), samʿ (12 at most), masmaʿ (2), 22. *earth* ʔarD (many) turaab (2/5C), 23. *eat* ʔakal (4½C), Ṭaʿim (2/3C), 24. *egg* none, bayD, djj, 25. *eye* ʔabsaar *eyes* (most but not by much) ʿayn *eye*,

[3] R. B. Lees, op. cit., 119.

[4] Morris Swadesh, Towards Greater Accuracy in Lexicostatistic Dating, IJAL 21.123.

[5] op. cit., 133–137.

[6] M. F. Barakat, al-Murshid ila Ayyat al-Qurʿan al-Karim wa Kalimatiha, al-Matbaʿah al-Hashimiyah, Damascus, 1939.

26. *fat* šuнuwm (1) dihn, samn, 27. *feather* none, riyš *splendid garments* (1), ziynah *adornment*, janaaн *crime*, *wing*, jild *skin*, 28. *fire* naar (most), нariyq (3), ɒaw?, ?abbuwliy, 29. *fish* none, samak, нawт, 30. *fly* тaar (2), saafar, nafar, 31. *foot* rijl (14), qadam (6), ?jj (2), ?rj, kff, 32. *full* ?i mtala? (1), muli? (2), mu'abbaa, matruws, 33. *give* ?a'тaa (10), ?dy (6), ?ahdaa, 34. *good* тayyib (35), нasan (17), zayn, kuwayyis, maliyн, maniyн, sнw, 35. *green* ?axɒar (occurs), 36. *hair* ?aš'aar (1), riyš *splendid garments* (1), 37. *hand* yad (4C), kff, 38. *head* ra?s (1C), 39. *hear* sami' (4C), tasamma', 40. *heart* qalb (4½C), fu?aad (3/4C), 41. *horn* none, qarn (0 except in name of Alexander the Great), ruwq, nafiyr, šabbuwr, buwq, 42. *I* ?ana (N), 43. *kill* qatal (most frequent), ðabaн *sacrifice* (occurs), 44. *knee* none, rukbah, kuw', zaawiyah, 45. *know* 'alim (13C), 'arif (1C), daraa (1C), 46. *leaf* waraqah (2), xuws, 47. *lie* ?i ttaka? *lean* (1), *recline* (1) munqa'ar *uprooted*, naam *sleep* (5), raqad *sleep* (1), lawaaqiн *rainbearing winds* (1), qa'ad *sit* (7), musannad *leaned against a wall* (1), ?i nsadaн, ?i nbaтaн, ?i ɒraja', ?i nɒaja', ?i stalqaa, tasaттaн, tamaddad, slqw, slqy, jɒ', 48. *liver* none, kbd, тuнaal, qasbah, sqwdaa?, 49. *long* тawiyl *continual* (1), *livelong* (1), madiyd, mustaтiyl, 50. *louse* none, gummal *tick* (1), jiraн duwdah

51. *man* rajul (54, including plural), ?insaan (64, singular only. Rejected without inspection of meaning. I understand it to mean *person*), bašar *mankind* (36), mar?, ?i mra? (10), rajjaal, maxluwq, ?aadamiy, zalmah, zamlah, ðalmah, ðamlah, 52. *many* kaθiyr (5C), waajid, 53. *meat* laнm (12), 54. *moon* qamar (1C), 55. *mountain* jabal (1 2/3C), ɒali', jaarib, jawb, 56. *mouth* faahu, ?afwaah (13), θamm, xašm, нalq, buwz, 57. *name* ?ism (4C), ?alqaab (1) 58. *neck* 'unuq (2), raqabah *captive*, 'itrah, 59. *new* jadiyd (6), нadiyθ (less), ?axiyr, 60. *night* layl (3¼C), 61. *nose* ?anf (2) xašm, minxaar, manxuwr, naxarah, naɒif, naðif, 62. *not* maa (N. selected as probably the most general of a number of negatives) 63. *one* waaнid (2C), 64. *person* ?insaan (64 counting singular only), rajul (54), bašar *mankind* (36), nafar (3), fulaan (1) ?aadamiy, šaxs, zawl, zamlah, zalmah, 65 *rain* maтar (6), šitaa? (1) ?aнyaa, нayaa' нaya, list, 66. *red* нumur (2) bunniy, 67 *road* sabiyl (7C), hudan (2 2/3C), siraa (1¼C), тariyq (5), тariyqah (2), тaraa?i (3), darb, rastah, mamarr, jaaddah, šaari' sikkah, xaтт, zqq, 68. *root* ?asl (2), saa *leg* (3), jiðʿ *trunk* (3), 'urq, jiðr, šurš, šluwš 69. *round* none, daa?irah *change* (3), mudaw wir, mudawwar, mustadiyr, muɒayğam kurawiy, murabba', 70. *sand* turaab (2/3C) raml, baтн, 71. *say* qaal (many), 72. *se* ra?aa (10½C), naðar (2½C), ?i тtala' (3) šaahad, šaaf, 'ayyan, taala', 73. *seed* нab *grain* (5), ðurriyyah *offspring* (1¼C), biðrah siyb, 74. *sit* qa'ad *sit* (6) *sit at home* (1) jalas, 75. *skin* jild, juluwd (9), bašarah

76. *sleep* nawm (3), naa?im (2), ruquw (1), 77. *small* sağiyr (8), 78. *smoke* duxxaa (2), duxnah, 'ajaaj, 79. *stand* qaam (1¼C waqaf (3), 80. *star* nujuwm *stars* (13) kawkab (6), 81. *stone* нijaar (11), нajar (10), saxarah (3), нasaah, нaswah, нayd, 82. *sun* šams (occurs), 83. *swim* none, 'wm, saбaн, šarra', ?i staнamm, 84. *tail* none, ðayl, ðanab liyyah, 85. *that* ðaalika (N), 86. *this* haaðaa (N), 87. *thou* ?anta (N), 88. *tongue* lisaan (1C) 89. *tooth* sinn (2), ɒirs, muqaddim 90. *tree* šajarah (1C), 91. *two* ?i θnayn (2/3C), 92. *walk* mašiy (20), saar *travel* (18), 93. *warm* (*hot*) нaamiy (3), нarr heat (3), mušawwab, нaarr, suxun, saaxin, 94 *water* maa? (2C), 95. *we* naнnu (N), 96 *what* maaðaa (N), 97. *white* ?abyaɒ (7) 98. *who* man (N), 99. *woman* ?i mra?ah (1C), нurmah, sayyidah, ?insaanah, 100. *yellow* safraa? (1)

The resulting Qoranic word-list contains a total of 88 items which I have compared with word-lists obtained from three Meccans at present living in eastern Saudi Arabia.

In eliciting the vocabulary which makes up the word-lists of the three informants Arabic was the language of communication.

The following elicitation techniques were used. (1) Obtained the noun desired by pointing. (2) Used pictures. (3) Acted out verbs where possible. (4) Used word-analogies such as "What on a bird corresponds to the hair on a man?" (5) Words were obtained through observation of the informant's speech. (6) Opposites were requested. (7) Definitions, such as "What is the animal which travels in the air?" were used.

In general, where the informant offered more than one word, I accepted the first if in my experience it was a word which was equivalent to the meaning I was seeking. At the beginning of the series of investigations, I asked for both classical and colloquial alternates. I found quickly that the meaning of "classical" varied radically from informant to informant, therefore I soon ceased to ask for both. Those words listed as classical alternates I have rejected. I also rejected words with a specialized meaning and words I inadvertently directly suggested to the informant.

After I had collected the three word-lists, I made up a composite word-list by including in it all words on which two or three contemporaries agreed. In one case, where no majority prevailed, I omitted the item.

One semantic item taken from the following word-lists will illustrate the method of presentation. The Qoranic equivalent, if it occurs, is always given first. The code-number (4) indicates the composite word-list defined above.

33. *give* a. ʔa ʿtaa (034), ʿataa (1), b. ʔahdaa (2) R. ʔaddaa (3) (2nd choice).

The four sources yielded four alternates divided into two mutually exclusive cognate-sets, a and b, and one rejected item, R. Cognate-set a consists of ʔa ʿtaa obtained from the Qoran (0) and informant 3. This item I have selected to represent the composite-word for contemporary Meccan as is indicated by (4). Informant 1 said ʿataa, cognate with ʔaʿtaa, and informant 2 said ʔahdaa, not cognate with the first two words and so belongs to cognate-set b. Informant 3 added ʔaddaa as an afterthought. I therefore rejected this item.

KEY TO COMPARATIVE LISTS: a, b, c, d: mutually exclusive cognate-sets. (0) Qoranic word-list. (1, 2, 3) word-lists of contemporary Meccan informants. (4) word-list of composite contemporary Meccan. R, rejected items.

1. *all* a. kull (0234) b. jamiyʿ (1), 2. *ashes* a. HUTaam (01) b. ramaad (2) rumaad (34) R. bagaaya *ashes of a cigaret* (1), xaraaba (1) third choice, 3. *bark*——(0), a. gišir (2) gišrah (14) b. liHaaʔ (3) 4. *belly* a. baTn (0) baTin (1234) R. kirš (3) 2nd choice, 5. *big* a. kabiyr (01234), 6. *bird* a. Tayr (01234), 7. *bite* a. ʿaDD (0134) b. yaDrus (2) R. yigazgiz (2) 2nd choice, 8. *black* a. ʔaswad (01234), 9. *blood* a. damm (01234), 10. *bone* a. ʿaðm (0) ʿaðmah (1) ʿaðum (234), 11. *breast* a. sadr (0124) sadur (3), 12. *burn* a. ʔi Htaraq (01) yiHrig (2) ʔa nHaraq (34), 13. *claw*——(0), a. ðafr (14) ðifir (3) b. ðalf (2), 14. *cloud* a. siHaab (01234) R. ğaym (3) 2nd choice, 15. *cold* a. baarid (01234) R. buraad (3) 2nd choice, 16. *come* a. ʔataa (0) b. jaaʔ (134) c. HaDar (2), 17. *die* a. maat (01234), 18. *dog* a. kalb (01234), 19. *drink* a. šarib (0) širib (1234), 20. *dry* a. yaabis (0) b. naašif (124) naššaf (3) R. gaaff (2) 2nd choice, 21. *ear* a. ʔuðun (0) ʔiðin (1234) R. widniy (3) 2nd choice, 22. *earth* a. ʔarD (01234), 23. *eat* a. ʔakal (01234), 24. *egg*——(0) a. bayD (3) bayDa (124), 25. *eye* a. ʔabsaar *eyes* (0) b. ʿayn (134) ʿaynu (2).

26. *fat* a. šuHuwm (0) šaHam (1234), 27. *feather*——(0) a. riyšah (134) riyš (2), 28. *fire* a. naar (01234) 29. *fish*——(0), a. samak (234) b. HowT (1) R. HowT (3) 2nd choice, 30. *fly* a. Taar (0134) yiTiyr (2), 31. *foot* a. rijl (0) b. qadam (2) gadam (134) R. rijil (1) 2nd choice, "colloquial and most common." 32. *full* a. ʔi mtalaʔ (0) malyaan (1) malʔ aan (2) ʔa tmallaʔ (34) R. šabʿaan *full of food* (1) 2nd choice, 33. *give* a. ʔaʿTaa (034) ʿaTaa (1) b. ʔahdaa (2) R. ʔaddaa (3) 2nd choice, 34. *good* a. Tayyib (01234), 35. *green*

a. ʔaxdar (01234), 36. *hair* a. ʔaš'aar (0) ša'ar (124) ša'ir (3), 37. *hand* a. yad (01234), 38. *head* a. ra'ʔs (0) ra'ʔas (2) raas (14) raas (3), 39. *hear* a. sami' (02) simi' (134), 40. *heart* a. qalb (01234), 41. *horn*——(0), a. qarn (134) garn (2), 42. *I* a. ʔana (01234), 43. *kill* a. quatal (01234), 44. *knee*——(0) a. rukbah (1234), 45. *know* a. 'alim (0) b. 'irif (134) yi'rif (2), 46. *leaf* a. waraqah (0) waraq (3) waragat šajarah (1) waraga (24), 47. *lie* a. ʔi ttaka' (0) b. tamaddad (234) c. ʔi nsadah (1) R. naam *sleep* (1), 48. *liver*——(0) a. kibid (1) kibda (24) kibdah (3), 49. *long* a. tawiyl (01234), 50. *louse*——(012), a. gamil (34),

51. *man* a. rajul (01) rajjaal (134) ragul (2), 52. *many* a. kaθiyr (0234) kθiyr (1), 53. *meat* a. lahm (0) laham (1234), 54. *moon* a. qamar (012) gamar (34), 55. *mountain* a. jabal (0134) gabal (2), 56. *mouth* a. ʔafwaah (0) famm (1234), 57. *name* a. ʔism (0124) ʔisim (3), 58. *neck* a. 'unuq (0) b. raqabah (2) ragabah (134), 59. *new* a. jadiyd (01234), 60. *night* a. layl (0234) layla (1), 61. *nose* a. ʔanf (0) b. xušm (1) xušum (234), R. ʔanf (1) 'classical', (2) 2nd choice, man . . . (1) First utterance, but not completed. When questioned, 1 said "manaaxiyr is Egyptian, understood in Mecca but not used." 62. *not* a. maa (01234), 63. *one* a. waahid (01234), 64. *person* a. ʔinsaan (0), b. šaxs (1234), 65. *rain* a. matar (01234), 66. *red* a. humur (0) ʔahmar (234) ʔahmir (1), 67. *road* a. sabiyl (0) b. tariyg (124) c. darb (3) R. sfilt *auto-road* (3) zugaag *alley* (3), 68. *root* a. ʔasl (0) b. jizir (1) c. saaq (2) d. 'irq (3)——(4), 69. *round*——(0) a. mudawwar (134) mustadiyr (2), 70 *sand* a. turaab (0) b. raml (1) ramal (24) c. batha (3) R. ramla (3) 2nd choice, 71. *say* a. qaal (0234) gaal (1), 72. *see* a. ra'ʔaa (0) b. šaaf (134) šuft (2) R. ra'ʔa (2) Said it was not colloquial, 73. *seed* a. habb (0) b. bizrah (1) c. dirra (24) dirrah (3), 74. *sit* a. qa'ad (0) ga'ad (3) b. jalas (14) galas (2), 75. *skin* a. jild (01234)

76. *sleep* a. nawm (0) naam (124) b. ragad (3) R. ragad (2) 2nd choice, 77. *small* a. saǧiyr (01234), 78. *smoke* a. duxxaan (01234), 79. *stand* a. qaam (0) b. wugif (1) wagaf (234), 80. *star* a. nujuwm (0) najimah (1) najim (2) najma (34), 81. *stone* a. hijaar (0) hujaar(3) hajar (124) R. suxuwr *big rock* (3), 82. *sun* a. šams (01234) 83. *swim*——(0) 'aam (134) ya'uwm (2), 84. *tail* ——(0) a. danab (1) b. ðayl (2) dayl (34) R. liyyah *tail of a sheep* (1) danab (3) 2nd choice, 85. *that* a. ðaalika (0) ðaak (3) daak (14) haadaak (2), 86. *this* a. haaðaa (0) haaða (13) haada (24), 87. *thou* a. ʔanta (02) ʔinta (34) ʔint (1), 88. *tongue* a. lisaan (01234), 89. *tooth* a. sinn (0124) sinna (3), 90. *tree* a. šajarah (0) šajara (1234), 91. *two* a. ʔi θnayn (0134) θnayn (2), 92. *walk* a. mašiy (0) mišiy (134) maša (2), 93. *warm* a. haamiy (0) b. haar (1234) R. saaxin (2) 2nd choice, 94. *water* a. maaʔ (02) mayya (134), 95. *we* a. nahnu (0) niḥna (2) ʔiḥna (134), 96. *what* a. maaðaa (0) maaza (2) b. ʔayš (134) R. ʔayš (2) 2nd choice, 97. *white* a. ʔabyad (01234), 98. *who* a. man (0) miyn (1234), 99. *woman* a. ʔi mra'ʔah (0) mara (2) b. hurmah (134), 100. *yellow* a. safraaʔ (0) asfar (1234)

The formula $\log r = \log C \div t$[7] is applied to obtain the rate-constant r. *t* is the time between Mohammed's death and 1958, 1.326 millenia. To support Swadesh's and Lees' estimates that "on the average about 81% of the basic-root-morphemes of a language will survive as cognates after 1000 years, for all languages, at all times," this rate-constant for Meccan Arabic should fall somewhere between the .8048 ± .0176 mill. set by Lees and the 6% above Lees' level set by Swadesh. In the comparison of each contemporary word-list with the Qoranic word-list the rate-constant r for Meccan Arabic is found to be .823, completely supporting Lees' and Swadesh's estimate. The data are tabulated as follows:

INFORMANT	COGNATES RETAINED	TOTAL NUMBER OF ITEMS COMPARED	C	r
1	68	88	.773	.823
2	68	88	.773	.823
3	68	88	.773	.823
4	68	87	.782	.831

[7] Swadesh, op. cit., 123.

This is an unexpectedly close result. As noted previously the method used in obtaining the Qoranic word-list would be expected to raise the value of the rate-constant. In addition the daily public recitation of the Qoran in mosques and schools throughout the Islamic world would lead one to expect that its vocabulary would have a higher retention-rate than almost any other vocabulary in the world. Such stability would tend to raise the value of the rate-constant still further above the mean found for other languages. Apparently these factors have not appreciably affected the rate-constant.

The rate of morphemic decay in Meccan Arabic seems to confirm most strongly the stability of the morpheme decay rate for all languages while failing to support "the common view about the conservatism of the Semitic languages" mentioned by A. I. Samarrai.[8]

DHAHRAN, SAUDI ARABIA

ANNOUNCEMENT

The Indiana University Graduate School's Committee on Uralic Studies announces establishment of the first American publication series in Uralic and Altaic studies.

The initial volume—AMERICAN STUDIES IN URALIC LINGUISTICS—will appear in March 1960. It will contain twelve contributions from as many specialists in this field teaching or trained in such American universities as Columbia, Indiana, Kansas, Michigan, Texas, and Washington. The Foreword is by President Frederick Burkhardt of the American Council of Learned Societies.

The Uralic and Altaic Series is edited by Thomas A. Sebeok, with Fred W. Householder, Jr., Felix J. Oinas, and Alo Raun serving as associate editors. Consulting editors are: John Lotz (Columbia University), Samuel E. Martin (Yale University), Nicholas N. Poppe (University of Washington), and Lewis V. Thomas (Princeton University).

Manuscripts of books, monographs, or extended articles dealing with any subject in the humanities or social sciences, in reference to any of the Uralic or Altaic peoples or languages, may be submitted to the editor at

Rayl House
Indiana University
Bloomington, Indiana, U. S. A.

Inquiries about subscriptions may be directed to the same address.

[8] A. I. Samarrai, Rate of Morphemic Decay in Arabic, IJAL 25.69.

RECENT PUBLICATIONS

General Linguistics

Belasco, Simon. The Differentiation of Allophonic and Non-allophonic Segments in Phonemic Analysis. American Speech 34.269–279, 1959.

Comparative Studies in Key-Words of Culture. Pilot Studies II, III. Edited by Gerhard Deeters and others. Bonn: Rheinische Friedrich - Wilhelms - Universität, 1959. Two volumes, xii, 155; v, 128.

Conklin, Harold C. Linguistic Play in its Cultural Setting. Lg 35.631–636, 1959.

Cowan, H. J. K. A Note on Statistical Methods in Comparative Linguistics. Lingua 8.233–246, 1959.

Fodor, I. The Origin of Grammatical Gender, I. Lingua 8.1–43, 1959.

Fodor, I. The Origin of Grammatical Gender, II. Lingua 8.186–214, 1959.

Francescato, G. A Case of Coexistence of Phonemic Systems. Lingua 8.78–86, 1959.

Gastil, Raymond P. Relative Linguistic Determinism. AL 1, no. 9, 24–38, 1959.

Hall, Robert A. Jr. Pidgin Languages. Scientific American 200, no. 2, Feb., 1959.

Hamp, Eric P. Graphemics and Paragraphemics. SIL 14.1–5, 1959.

Hiorth, F. Origin and Control of Meaning Hypotheses. Lingua 8.264–293, 1959.

Hoenigswald, Henry. Some Uses of Nothing. Lg 35.409–420, 1959.

Householder, Fred W. Jr. On Linguistic Primes. Word 15.231–239, 1959.

Jakobson, Roman. Boas' View of Grammatical Meaning. In: The Anthropology of Franz Boas, ed. by Walter Goldschmidt. Mem. 89, AAA, 1959.

Kaper, W. Kindesprachforschung mit Hilfe des Kindes. Groningen: J. B. Wolters 1959. xxiii, 244 pp.

Kirk, Paul L. A Note on Morphemic Fusion and Empty Morphs. SIL 14.7–9, 1959.

Luria, A. R. The Directive Function of Speech, I: Its Development in Early Childhood. Word 15.341–352, 1959.

Malkiel, Y. Studies in Irreversible Binomials. Lingua 8.113–160, 1959.

Mol, H. and E. M. Uhlenbeck. Hearing and the Concept of the Phoneme. Lingua 8.161–185, 1959.

Murdock, George P. Cross-Language Parallels in Parental Kin Terms. AL 1, no. 9, 1–5, 1959.

Olmsted, D. L. The Identity Theorem. Gen. Lgstics. 4.38–39, 1959.

Pike, Kenneth L. Language in Relation to a Unified Theory of the Structure of Human Behavior, Part III (Chapters 11–17). Glendale, Calif.: Summer Inst. of Linguistics. Prelim. Ed., 1960. 146 pp.

Siertsema, B. Problems of Phonemic Interpretation II. Long Vowels in a Tone Language. Lingua 8.42–64, 1959.

Stockwell, Robert P. Structural Dialectology: a Proposal. American Speech 34.258–268, 1959.

Uldall, H. J. On the Preparation of a Text. Archivum Linguisticum 11.1–17, 1959.

Voegelin, C. F. and F. M. Guide for Transcribing Unwritten Languages in Field Work. AL 1, no. 6, 1959.

Welmers, William E. Tonemics, Morphophonemics, and Tonal Morphemes. Gen. Lgstics 4.1–9, 1959.

Wolff, Hans. Subsystem Typologies and Area Linguistics. AL 1, no. 7, 1–8, 1959.

American Indian Languages

Bascom, Burt. Tonomechanics of Northern Tepehuan. Phonetica 4. 71–88, 1959.

Baumhoff, Martin A. California Athabascan Groups. Anthro. Records (Univ. of California) 16.157–231, 1958.

Beasley, David, and Kenneth L. Pike. Notes on Huambisa Phonemics [Huambisa, an Amerindian language of Peru.] Lingua Posnaniensis 6.1–8, 1957.

Bergsland, Knut. Aleut Dialects of Atka and Attu. Trans., American Phil. Soc., n.s., v. 49, part 3, 1959. 128 pp.

Bibliographical Bulletin of American Anthropology. Vols. 19–20, 1956–57. Mexico, D. F.: Inst. Panamericano de Geografia e Historia, 1959.

Bright, William and Elizabeth. Spanish Words in Patwin. [An Amerindian language of California.] Romance Philology 13.161–164, 1959.

Bright, William, and D. L. Olmsted. A Shasta Vocabulary [Shasta, an Amerindian language of California.] Papers, Kroeber Anthro. Soc., no. 20, 1–55, 1959.

Chafe, Wallace L. The Classification of Morphs in Seneca. AL 1, no. 5, 1–6, 1959.

Chafe, Wallace L. Internal Reconstruction in Seneca. Lg 35.477–495, 1959.

Clark, Lawrence E. Phoneme Classes in Sayula Popoluca [An Amerindian language of Mexico] SIL 14.25–33, 1959.

Fernandez de Miranda, Maria Teresa. Fonemica del Ixcateco [of Oaxaca, Mexico]. Pub. 3, Inst. Nacional de Antropologia e Historia, Mexico, D. F., 1959. 21 pp.

Fraenkel, Gerd. Yaqui Phonemics. AL 1, no. 5, 7–18, 1959.

Holland, William R. Dialect Variations of the Mixtec and Cuicatec Areas of Oaxaca, Mexico. AL 1, no. 8, 25–31, 1959.

Hung, Beverly Y. P. On the Phonemic Status of Navaho Stress. AL 1, no. 9, 20–23, 1959.

Kroeber, A. L. Northern Yokuts. AL 1, no. 8, 1–19, 1959.

Landar, Herbert J. The Navaho Intonational System. AL 1, no. 9, 11–19, 1959.

Olmsted, D. L. Palaihnihan and Shasta III: Dorsal Stops. Lg 35.637–644, 1959.

Pickett, Velma, and Others. Castellano-Zapateco; Zapateco-Castellano. Mexico, D. F.: 1959. (Publicado por el Inst. Linguistico de Verano en cooperacion con la Dirección general de Asuntos Indigenas de la Secretaria Educacion Publica). 246 pp.

Salzman, Zdenek. Araphaho Kinship Terms and Two Related Ethnolinguistic Observations. AL 1, no. 9, 6–10, 1959.

Swadesh, M. Indian Linguistic Groups of Mexico. Mexico, D. F.: Escuela Nacional de Antropologia e Historia, 1959. 14 pp.

Swadesh, Mauricio. Mapas de Classificación Lingüistica de México y las Américas. Cuadernos del Inst. de Historia (Mexico, D. F.), Serie Antropologica no. 8, 1959.

Taylor, D. La catégorie du genre en caraïbe insulaire. Bull. de la Soc. de Linguistique de Paris 54.201–207, 1959.

Miscellaneous

Anthropological Research in Netherlands New Guinea since 1950. By the Bureau for Native Affairs, Hollandia, Netherlands New Guinea. Oceania Monographs, no. 10 (reprinted from Oceania 29.132–163), 1959.

Cense, A. A., and E. M. Uhlenbeck. Critical Survey of Studies on the Languages of Borneo. 'S-Gravenhage: Martinus Nijhoff, 1958. 82 pp.

Linguistics and the Teaching of English as a Foreign Language. Special Issue, Language Learning, June, 1958. 136 pp.

Morgan, Ralph Jr. The Lexicon of Saint Martin Creole. AL 2, no. 1, 7–29, 1960.

Morgan, Ralph Jr. The Phonology of Saint Martin Creole. AL 1, no. 8, 20–24, 1959.

Wallis, Ethel E., and Mary A. Bennett. Two Thousand Tongues to Go, the Story of the Wycliffe Bible Translators. New York: Harper and Bros., 1959. ix, 308 pp.

* * * * *

Reprints of the six parts of Gladys Reichard's "A Comparison of Five Salish Languages" (IJAL 24 No. 4, 25 No. 1–4 and 26 No. 1) have been bound and are available for $1.00. Orders may be sent to the Business Manager of IJAL.

PART III

International Journal of American Linguistics

Volume 26 Number 3

July 1960

Publication Thirteen
of the
INDIANA UNIVERSITY RESEARCH CENTER
in
ANTHROPOLOGY, FOLKLORE, AND LINGUISTICS

LINGUISTIC DIVERSITY IN SOUTH ASIA
Studies in Regional, Social and Functional Variation

Edited with an introduction
By Charles A. Ferguson and John J. Gumperz

Second-class postage paid at Baltimore 2, Md.

LINGUISTIC DIVERSITY IN SOUTH ASIA

Studies in Regional, Social and Functional Variation

LINGUISTIC DIVERSITY IN SOUTH ASIA

Studies in Regional, Social and Functional Variation

Edited

With an Introduction

by

Charles A. Ferguson and John J. Gumperz

INDIANA UNIVERSITY
RESEARCH CENTER IN
ANTHROPOLOGY, FOLKLORE, AND LINGUISTICS

Editor: C. F. Voegelin

Director of Publications: Thomas A. Sebeok

Publication Thirteen, pp. viii + 118, Issued July 1960
Also
Part III of the
International Journal of American Linguistics
Vol. 26, No. 3 (1960)

Priced separately, $3.50. Orders will be received by the
Director of Publications of the Research Center, at
Indiana University, Bloomington, Indiana

FOREWORD

The present volume is the outgrowth of a panel discussion "Language and Culture Dynamics in South Asia" which was held at the annual meeting of the American Anthropological Association in Chicago, December 1957. That symposium was sponsored by the Committee on South Asian Languages of the Association for Asian Studies and was organized by Charles A. Ferguson, who was at that time Chairman of the Committee. In the unavoidable absence of Professor W. Norman Brown of the University of Pennsylvania, who had hoped to preside, Ferguson also acted as chairman of the panel.

At that session papers by Bright, Chowdhury, and Dimock were presented by the authors essentially in the form in which they appear in this volume. In addition Professor Uriel Weinreich of Columbia University spoke on "The Troubles of Hindi" and McCormack's paper, which he submitted from India, was read by Bright. John J. Gumperz, the succeeding Chairman of the Committee on South Asian Languages, and Selig S. Harrison, Associate Editor of the New Republic, acted as discussants.

Before publication of the papers, Weinreich's study was withdrawn as being of somewhat different approach and more appropriate for publication elsewhere, and relevant articles by Pillai and by Naim and Gumperz were added to the group.

At the invitation of Professor Thomas A. Sebeok of Indiana University the Committee agreed to publish the volume in the format of a Memoir of the International Journal of American Linguistics (IJAL), with additional distribution as a supplement to the journal Indian Linguistics.

Ferguson and Gumperz undertook the editing of the papers. They made some attempt to bring the transcription, section numbering, and style of the various authors closer together without interfering with the individual preferences of the authors; in several cases they shortened the papers; and they provided the Introduction to the volume. They would like to

express here their appreciation of the patience and helpfulness on the part of authors and publishers and to apologize for the long period of time which the editing process took.

Charles A. Ferguson

John J. Gumperz

CONTENTS

Foreword v

Introduction ..... Charles A. Ferguson
John J. Gumperz 1

Linguistic Change in Some Indian Caste
Dialects ..... William Bright 19

Tamil — Literary and Colloquial ... M. Shanmugam Pillai 27

Literary and Colloquial Bengali in Modern
Bengali Prose ..... Edward C. Dimock 43

The Language Problem in East
Pakistan ..... Munier Chowdhury 64

Social Dialects in Dharwar
Kannada ..... William McCormack 79

Formal and Informal Standards in the Hindi
Regional Language Area ..... John J. Gumperz
C. M. Naim 92

INTRODUCTION

Charles A. Ferguson

Center for Applied Linguistics

John J. Gumperz

University of California, Berkeley

0. Introductory

1. Variety, language, dialect

2. Geographical dialects

3. Social dialects

4. Style and superposed varieties

0. The diversity of languages in South Asia and some of its implications are often noted,[1] but relatively little attention is paid to the dialect diversity which exists within the individual languages in the area. All the major languages of India, Pakistan, and Ceylon have regional dialects as well as dialects which correlate with differences of social position, and they also have special literary forms of language which differ to a greater or lesser degree from the ordinary colloquial. Such variation in language is a widespread phenomenon in the world — it might almost be called the 'normal' state of languages — but the exact nature of the dialect diversity in South Asia is not well understood, and its relevance to non-linguistic aspects of South Asian society and culture has never been systematically explored. The studies in this volume constitute a first attempt to describe and interpret the facts of dialect diversity in several specific languages of this part of the world. No great effort is made to carry the interpretation far afield from linguistics, but each of the studies contains suggestive material for the approaches of other disciplines in the study of contemporary South Asia. Since these studies

are addressed both to linguists and to non-linguists, the editors have felt it necessary to provide in this introduction a brief summary of the theoretical framework which linguistic science currently offers for the treatment of intra-language variability.

1. First of all, one of the fundamental problems of linguistics is the delimitation of languages as the 'natural' units of linguistic analysis and classification. Put more directly, one of the concerns of linguistics is to find a consistent, systematic way to answer such questions as: Are these two particular varieties of human speech one language or two? Questions of this kind arise all over the world. Are British English and American English one language or two? Do all the Arabic dialects constitute one language or several? Familiar examples from South Asia include: Is Hindi one language or several? Are Konkani and Marathi two languages or one? How many Dravidian languages are there?[2] Questions like this cannot be answered until some kind of definition of language is widely accepted as valid, or possibly until the concept of language as a unit is rejected and another formulation of the questions becomes accepted.

Surprisingly little attention is paid to this problem, however, and most definitions of language which are in vogue among linguists today are more concerned either with setting off speech behavior from other human activity or with setting off linguistic systems from other semiotic systems than they are with defining the limits of single languages. For example, linguists frequently define language as "a system of arbitrary vocal symbols by means of which a social group cooperates."[3] This definition offers no clues whatever for the solution of the problem since two social groups may use the same language or the same social group may use two or more languages; and in any case the definition fails to specify 'social group' either in linguistic or non-linguistic terms. Accordingly, the best that can be done here is to summarize in a relatively crude, cover-all definition the uses of the term 'language' in this sense by reputable linguists.[4]

In order to do this, one additional concept requires definition, the minimal unit which serves as the 'normal' object

of linguistic description. A whole language is often too varied to be readily analyzed by current descriptive techniques, and descriptive linguists generally restrict their object of study with considerable care. Harris[5] says, "The universe of discourse for a descriptive linguistic investigation is a single language or dialect." Bloch[6] uses the term 'idiolect,' the "totality of the possible utterances of one speaker at one time in using the language to interact with one speaker." These brief definitions, as the authors of them are fully aware, are not completely satisfactory. The idiolect concept for example is a necessary one, since in practice languages are always studied in terms of the speech of single speakers.[7] As a unit of analysis however an idiolect is defined by extra-linguistic criteria, and homogeneity of structure is not a necessary requirement. Thus the speech of two individuals may be so very similar as to cause no difficulty in description while two styles of speech of one individual may be very different, and no one has shown how to delimit styles rigorously in the speech of one person or of many.

At the present state of linguistics it is probably wise to sidestep the issues involved in devising a precise operational definition and to settle for a suggestive description. There is no term in regular use for this concept of the object of descriptive linguistic analysis, but the word 'variety' has been used by several authors in this sense and will be so used here.[8] A variety is any body of human speech patterns which is sufficiently homogeneous to be analyzed by available techniques of synchronic description and which has a sufficiently large repertory of elements and their arrangements or processes with broad enough semantic scope to function in all normal contexts of communication. With this concept clarified we may now attempt to define a language.

A first approximation to the definition may be in terms of mutual intelligibility. Other considerations aside, two varieties of human speech are said to be the same language if each is readily intelligible to speakers of the other. The difficulties of a definition of this kind are many, however, and are reminiscent of the difficulties of the biologist's use of the possibility of interbreeding as the primary criterion in the definition of the species.

In the first place mutual intelligibility is difficult to determine[9] (or even impossible, as with dead languages). Second, there seems to be little direct correlation between similarity of phonological or grammatical structure on the one hand and mutual intelligibility on the other. This lack of correlation is disturbing because linguists generally regard these structures as fundamental in human speech and often prefer to describe linguistic relationships in these 'internal' terms. Also, we frequently find varieties of speech which do not show high mutual intelligibility but which on totally different grounds we wish to regard as the same language, and vice versa. A well-known example is found in the varieties at the extremes of the German-speaking area which are quite different, with relatively low mutual intelligibility. They are connected, however, by a series of intervening varieties which provide a gradual transition, and linguists generally agree in regarding all as varieties of one language.

On the other hand, certain varieties are regarded as different languages in spite of their high degree of mutual intelligibility. For example, there is a gradual transition between varieties of Dutch and German, but because of the existence of two relatively stabilized, widely-used standard varieties, Dutch (ABN) and German (Schriftdeutsch), two speech communities and hence two languages, are usually recognized; any transitional variety is assigned to the language whose standard form is accepted in the community, regardless of the internal linguistic similarities.

Another commonly used criterion for differentiating languages is that of difference in historical development. Two sets of varieties are said to constitute different languages if they differ significantly in their treatment of certain phonological and morphological features of a reconstructed parent variety. In contrast to the problems encountered in the determination of mutual intelligibility, historical relationship is relatively easy to establish by the rigorous and generally accepted methods of historical reconstruction. Grierson used this criterion in classifying speech distribution in what is now called the Hindi regional language area. He grouped the local varieties into five groups, which he calls languages: Bihari, Eastern Hindi, Western Hindi, Rajasthani, and Pahari.

Each group is set off from the others by peculiarities of grammar and pronunciation. Modern Hindi-Urdu is part of the Western Hindi group, but its historical relationship to the other local varieties is by no means always very close. Bihari is more directly related to Bengali, the regional language of Bengal; some forms of Rajasthani are usually classed with Gujerati to the south and the Pahari dialects are closest to Nepali, the language of Nepal.[10]

Classifications of the above kind are essential in historical linguistics, and they provide valuable information for students of culture history. They do not, however, present all the facts of language distribution. In spite of the diversity of local speech forms and their relationship to neighboring regional languages in Grierson's area, Hindi-Urdu is recognized as the language of literature and government throughout. A large proportion of the modern urban population and of the educated speaks it and in other ways regards itself as belonging to a single Hindi-Urdu speech community. Many modern writers thus disregard the historical relationships and refer to all local speech forms as varieties of Hindi.[11] As in the case of the German-Dutch area, the presence of a superposed variety, such as a literary standard, is an important factor in the determination of language boundaries. This factor is irrelevant, however, if the standard language in use in an area is extremely divergent or obviously unrelated to the variety to be classified. All this brings us to our working definition.

A language consists of all varieties (whether only a single variety or an indefinitely large number of them) which share a single superposed variety (such as a literary standard) having substantial similarity in phonology and grammar with the included varieties or which are either mutually intelligible or are connected by a series of mutually intelligible varieties.

It is evident that, as with the term idiolect, purely linguistic criteria are not currently accepted as sufficient for the delimitation of a language. Rather, a given set of varieties must meet certain minimum linguistic conditions (e. g. structural similarity within the set, structural difference from varieties outside the set) as well as certain sociological conditions (e. g. use of standard, speakers' feeling of belonging

to speech community) in order to be regarded as a single language.

When the concepts of 'variety' and 'language' have been explained it is feasible to define the linguistic notion of 'dialect,' which is crucial to the present studies. It must be made clear at the outset that two conventional uses of this term are to be disregarded. One is the use of 'dialect' to mean an inferior language of some kind. This use of the term is still widespread in such expressions as: "Mr. A speaks six languages and ten dialects." "The inhabitant of that country speaks dozens of languages and innumerable dialects." Since no linguistic criterion has ever been devised to differentiate superior and inferior languages, linguists prefer to operate without this distinction.[12] The other is the use of 'dialect' to mean any non-standard variety of a language. This use is still current in expressions such as: "He doesn't speak good German; he speaks only dialect." "These children are learning to speak the language without a trace of dialect." "This is a dialect word not preserved in the literary language." Many languages have one variety which is in certain respects dominant, the 'standard' variety, and it is often useful to distinguish between the standard variety and non-standard varieties of a given language, but linguists have found it more useful to leave the word 'dialect' colorless and equally applicable to standard and non-standard varieties. Thus linguists may refer to the 'standard dialect' of a given language, and even if they continue to use the traditional expression 'standard language' they generally regard this as a dialect, often, of course, the most important from some points of view.[13]

In the terminology of linguistics a dialect is, roughly speaking, something between a variety and a language. Sometimes in the analysis of a language an obvious classification of varieties into dialects is clear to the linguist. More often a number of alternative analyses are possible, and any particular classification into dialects runs counter to at least some features which could be used for classification. The elementary concept of dialectology is the isogloss. Any boundary which constitutes the limits of use of a particular linguistic feature is called an isogloss. For example, the word <u>frappe</u> is regularly used in the area around Boston to

refer to a drink consisting of milk, ice cream, and flavoring beaten together. In most of the United States this drink is called a milk shake, a term which in the Boston area refers to a similar drink made without ice cream. The boundary which shows the limits of the use of frappe in this sense, and which could be plotted on a map, is called an isogloss. In addition to isoglosses concerned with the occurrence of words in certain meanings (lexical isoglosses), linguists also make use of phonological and grammatical isoglosses.

In classifying dialects linguists generally prefer to place emphasis on isoglosses (a) which cut boldly across large areas as opposed to those which are relatively local, or (b) which bundle together as opposed to those which are relatively isolated, or (c) which correlate with non-linguistic criteria of classification such as differences in material culture, religion, political units, and the like.

Since dialect classification is largely arbitrary it might seem to be of relatively little value or even misleading, but in fact the delimitation of dialects is of great importance, both within linguistics proper and for extra-linguistic fields. In linguistic theory, a dialect may be regarded as the beginning of a linguistic split, as a step in linguistic differentiation. In other words a dialect is a potential new language, and the concept is comparable in validity and significance to that of the 'sub-species' or 'variety' of biology. For the extra-linguistic importance of the dialect we need only note its obvious relevance to the units of other social sciences such as geographical region, social class, or role.

A dialect is any set of one or more varieties of a language which share at least one feature or combination of features setting them apart from other varieties of the language, and which may appropriately be treated as a unit on linguistic or non-linguistic grounds. Because of the arbitrariness of this concept, linguists using the term 'dialect' for a particular language generally feel under obligation to explain and justify the criteria used for their classification in that language.[14]

2. Geographical dialects. The extent and nature of dialect differentiation in a language may be described in terms of two major correlates: density of communication and

inter-speaker attitudes. It cannot be maintained that these are the only relevant factors since the significance of such matters as the structure of the language (Sapir's 'drift'),[15] interference of other languages, phonetic symbolism, or technological change can be shown, but on the whole these two correlates provide a satisfactory frame of reference for the analysis of dialect diversity.

The assumption with regard to density of communication may be phrased in general terms as follows: Other things being equal, the more frequently speakers A and B of language X communicate with each other by means of X, the more the varieties of X spoken by them will tend to become identical.[16] It follows from this that isoglosses will tend to coincide with breaks or lines of weakness in communication.[17] Since such breaks or lines of weakness will often correlate with natural physical boundaries such as mountains or rivers, or with political boundaries, or with economic limits (on the flow of goods and services), it is clear that a large proportion of dialect diversity will be subject to mapping. Possibly because of the effectiveness of cartographic representation of dialect differences, this has been one of the most highly developed fields of work in dialectology. Dialect atlases have been published for many languages, and many special techniques of mapping have been developed. Apart from limited special studies most books in dialectology are either dialect atlases or grammars of non-standard varieties.

In South Asia, surprisingly enough, there has been very little work of this kind. In the monumental Linguistic Survey of India Grierson offered sample texts and brief descriptions of many dialects and in some cases provided maps indicating what he believed to be the boundaries of major dialects, but he did not attempt dialect atlases.[18] Almost all the work in South Asian dialectology published since Grierson has been based on his work, with only a very meager amount of original investigation, and Grierson's volumes remain by far the most important source of data for the social scientist concerned with the distribution and dialect diversity of South Asian languages.[19]

None of the studies included in this volume is directed primarily to the analysis of mappable features of dialect difference although all of them assume a background of regional

dialects. Accordingly, there is no need here for further discussion of the concepts of dialectology concerned with shape and distribution of regional isoglosses, such as 'relic area,' 'fan,' 'bundle,' 'transition zone,' 'focus.' The important thing is to note that all the techniques elaborated in the production of dialect atlases elsewhere in the world are available for use in the study of South Asian dialectology, and we may express the hope that significant studies in this field will be forthcoming in the next few years.

3. Social Dialects. The assumption with regards to inter-speaker attitudes may be stated in terms of two processes. First: any group of speakers of language X which regards itself as a close social unit will tend to express its group solidarity by favoring those linguistic innovations which set it apart from other speakers of X who are not part of the group. The existence within a speech community of social distinctions such as those of caste, class, professional guild therefore gives rise to differential rates of linguistic change, favoring the creation of new speech differences or the preservation of existing ones. On the other hand: other things being equal, if two speakers A and B of language X communicate in language X and if A regards B as having more prestige than himself and aspires to equal B's status, then the variety of X spoken by A will tend towards identity with that spoken by B. It may seem that this is a circular statement since one of the indications of the attitude of prestige is linguistic imitation but it is assumed that prestige relationships may be measured independently of language. As Leach puts it in describing a somewhat different multi-lingual situation: "For a man to speak one language rather than another is a ritual act, it is a statement about one's personal status: to speak the same language as one's neighbor expresses solidarity with those neighbors, to speak a different language from one's neighbors expresses social distance or even hostility."[20]

The above two processes create a condition of equilibrium:[21] as old speech differences vanish, new ones arise with the addition of population elements from the outside or through the formation of fresh social distinctions. The total range of speech diversity within a speech community, i.e. the actual language distance between all the varieties found there, is a

function of the density of intergroup communication. In societies like our modern American one where social class is fluid and mass communication media are highly developed and shared by most, we would expect the range of variation to be relatively small. Differences would most frequently appear on the level of phonetics or the lexicon but rarely affect the fundamental phonological or morphological structure of the language. In Asia, however, where intergroup communication is severely limited by ritual restrictions we would expect these differences to be much greater.

The parallels between speech variation and social stratification are manifold. This field of investigation has been worked very little by the dialectologist, however, whether because of the difficulty of presenting the data, or the problem of identification of social groupings, or simple historical accident in the development of the science. A handful of valuable studies in this field do exist for various languages and show great promise for the development of sociolinguistic analysis,[22] but until now very few such studies have appeared dealing with South Asian languages.[23]

Two of the papers in this volume have direct bearing on the problem of social dialects. Bright describes the speech of a Brahmin and a non-Brahmin speaker of the Mysore dialect of Kannada. Both are members of the western educated class and have had college education; the linguistic gap between them is nevertheless considerable. Of particular interest is the difference in receptiveness to outside innovation. The non-Brahmin dialect has made a number of changes in the Old Kannada grammar and sound system but shows little influence from outside borrowings. The Brahmin dialect on the other hand, possibly because of the influence of the literary language, has been much slower to change the native Kannada patterns while it has been profoundly affected by borrowings from Sanskrit and English. McCormack in studying the Dharwar dialect of Kannada finds three basic speech strata corresponding to the division between Brahmin, touchable non-Brahmin and untouchable castes. He also discusses the function of dialect differences as clues for the determination of a speaker's place in the caste hierarchy, a problem which has received considerable attention lately.[24]

The significance of the linguistic studies for broader research is clear. For example the problem of quantifying caste difference in terms of behavior and attitudes is a central problem in the analysis of South Asian social structure.[25] The limited amount of data we have already points to some interesting parallels. If we compare the dialect distribution described by Bright and McCormack and before them by Bloch and Aiyar with similar information on North India provided by Gumperz, we find that the southern speech communities (or at least those in the Kannada, Tamil and Tulu areas) show a fundamental gap between Brahmin and non-Brahmin speech. In the North on the other hand (or at least in the Western Hindi and Punjabi and probably also in the Bengal area) Brahmin dialects do not seem to differ significantly from those spoken by other high caste touchables. This bears out other recent independent anthropological findings regarding the difference in status between northern and southern Brahmins.

The analysis of linguistic behavior in the context of social science offers a new dimension of measurement and one that is particularly accessible and quantifiable compared to many other aspects of behavior. There is every possibility that more detailed and better planned sociolinguistic investigation of this kind in South Asia will lead the way into a whole new area of the analysis of society.

4. Style and superposed varieties. The third kind of variation to be described here is different in nature from the preceding two. Geographical or social dialects tend to be mutually exclusive in the sense that the normal individual is a native speaker of only one dialect, and if he speaks more than one this is a sign of geographical or social transition of some kind on his part. But there are also linguistic variations which regularly coexist in the speech of individuals, with their use reflecting some kind of situational or role differences.

It is part of the nature of language, or of human culture in general, that there should be individual differences in performance and differences in the performance of a given individual in various situations. For example, a speaker normally has 'favorite expressions' (e.g. words, phrases, constructions) which he uses more frequently than other speakers; and any adult speaker normally uses different kinds of speech talking

to his young children, participating in a religious ceremony, chatting with old friends, and explaining something to a stranger. To take somewhat different examples, in English the two words light and illumination have roughly the same meaning, but light tends to be used in less formal, less pretentious situations. Or, again, the words fore, aft, deck, companionway have meanings very similar to those of front, back, floor, stairway, but the first set is generally used either in a maritime setting or to suggest it, or by people who regularly live and work in such a setting.

All languages seem to show variations of this kind. As is the case with social dialects, however, the exact range of these variations and the level of linguistic structure at which they appear vary significantly. In America we find only minor differences in lexicon, syntax, stress and intonation, which do not materially complicate the task of linguistic analysis. Other parts of the world however show more fundamental variations. One of the most extreme situations of this type is that described by Garvin and Riesenberg on Ponape, where different sets of bound morpheme alternants and lexical items were used in social situations calling for honorific and non-honorific speech. Whenever two styles have striking morphological variation (e. g. case endings vs. no case endings) or extensive lexical pairing (e. g. pairs of different words for common items such as 'head,' 'go,' 'now'), or when both styles in question are regularly used under appropriate circumstances by a large segment of the speech community (e. g. all educated speakers of the language), then special language situations must be recognized which deserve study and which clearly present theoretical and practical problems of quite a different nature. The investigation of stylistic differences of this kind is of interest to the linguist since they constitute a fundamental mechanism of language and their analysis provides an insight into more general aspects of speech dynamics. For other social scientists correlation of style with social situation, personality, and similar factors offers valuable clues to the understanding of non-linguistic human behavior.[26]

Some styles of speech are sufficiently striking and extensive to require study as more or less autonomous marginal or partial systems within a language,[27] and often enough the styles

appropriate to certain occasions are full-fledged varieties of the language existing over and above the primary or 'natural' speech of individuals. There are many kinds of these superposed varieties, the most important being the various types of standard languages.

Very often one variety of a language, e.g. a certain local dialect, becomes widely accepted as standard and is learned by speakers of other varieties in addition to their native variety. Among typical examples of this standard language vs. local dialect situation we may cite important languages of Europe and Asia (e.g. German, Italian, Persian) where many speakers use a local dialect in familiar or 'home' situations but use the standard language in more formal or 'outside' situations. In some speech communities there may even be several layers of superposed varieties of this sort, with national and regional standards in addition to local dialects. To complicate this picture there may also be regionally standardized pronunciations of the national standard as well as various mixed varieties appropriate to certain occasions or used by individuals who have not fully mastered certain levels.[28] This layer phenomenon is fairly frequent in South Asia and has recently been described for Hindi.[29] Another kind of superposed variety is illustrated by the classical-colloquial dichotomy of Arabic, modern Greek, or Swiss German where the standard form of the language is used for written and formal spoken purposes only, the other for ordinary conversation.[30] The spread of a particular standard variety is determined by social and political factors and often has little relation to the actual language distance between the standard and the local dialects. Thus, returning to our German example, we find High German used as a standard by many speakers whose native dialects are much closer to Dutch. Another interesting case in point is that of Southern Nepal where native speakers of Maithili, a dialect related to Bengali, are advocating the adoption of Hindi as a standard language. Other Maithili speakers across the border in Bihar are opposed to Hindi and favor the creation of a new literary standard based on medieval literary Maithili.

Stylistic variation and superposition are important factors throughout South Asian society and four studies in this collection are explicitly concerned with this kind of variation: Pillai's

study of Tamil deals with a situation which is quite similar to the above mentioned cases of Arabic, modern Greek, and Swiss German. There are two distinct literary and colloquial styles, showing striking differences in phonology and morphology. The former serves as the sole medium of writing and formal speech making, while the latter is used for conversation among the educated. Dimock deals with the two standard varieties of Bengali, where conditions are somewhat different. Sadhu Bhasa, like literary Tamil, is employed for writing and formal speeches. Colit Bhasa, the conversational medium of the educated, however, has become accepted as a vehicle of literature in its own right during the last seventy five years. It serves as the literary symbol of the newly arising urban middle class, in contrast to traditional culture represented by Sadhu Bhasa. Chowdhury describes the relationship between the two standards and the regional dialects of East Bengal and details the social situations in which they are used. He also points to the incipient development of a third Eastern Bengali standard variety, resulting from political conditions arising out of the partition of Bengal. The Naim-Gumperz paper deals with the Hindi-Urdu area where two literary varieties have developed from a single conversational standard and have become symbols of the competing Hindu-revivalist and Indo-Muslim cultural complexes.

Notes

1. Cf. S. Harrison, The challenge to Indian nationalism, Foreign Affairs 34. 620-36 (1956); M. Windmiller, Linguistic regionalism in India, Pacific Affairs 27. 291-318 (1954); R. Pieris, Bilingualism and cultural marginality, British Journal of Sociology 2. 328-39 (1951). The fullest factual account of South Asian languages is still G. Grierson, Linguistic survey of India (19 vols. Calcutta, 1903-28). A good summary is provided in S. K. Chatterji, Languages and the linguistic problem (London, 1943). An extensive bibliography on politico-linguistic problems of South Asia is found in S. Harrison, The dangerous decades (New York, 1955). S. K. Chatterji, Indo-Aryan and Hindi (Ahmedabad, 1942) is a valuable historical treatment of one important area.

2. For a discussion of the widely divergent statistics on such questions in South Asia, see J. J. Gumperz, Some remarks on regional and social language differences in India, in

M. Singer (ed.), Introducing India in liberal education (Chicago, 1957).

3. Cf. B. Bloch and G. L. Trager, Outline of linguistic analysis (Baltimore, 1942).

4. For valuable recent discussions of the whole language-dialect question, see A. Martinet, Dialect, Romance Philology 8.1 (1954) and C. F. Hockett, A course in modern linguistics Chs. 38, 39 (New York, 1958). Perhaps the best discussion of this problem in a broader sociological framework is that of H. Kloss, Die Entwicklung neuer germanischer Kultursprachen (Munich, 1952).

5. Z. S. Harris, Methods in structural linguistics 9 (Chicago, 1951).

6. B. Bloch, A set of postulates for phonemic analysis, Language 24.7 (1948). Other authors define 'idiolect' differently. Cf. Hockett op. cit.

7. Cf. Hockett, op. cit., 321.

8. A. Kelkar in unpublished works has used for this the term 'synlect' based on the term 'lect' which is a key item in his comprehensive terminology for linguistics.

9. Some first steps have been taken in this field, and as better testing techniques and more satisfying methods of quantification are developed, measures of mutual intelligibility may become increasingly useful for linguistic descriptions. Cf. C. F. Voegelin and Z. S. Harris, Methods for determining intelligibility among dialects of natural languages, Proceedings of the American Philosophical Society 95. 322-9 (1951); J. E. Pierce, Dialect distance testing in Algonquian IJAL 18. 203-10 (1952).

10. Grierson, op. cit., vol. I. The details of Grierson's classification are open to question on technical grounds but this does not invalidate the historical approach as such.

11. See for example Dhirendra Varma, Hindi literature, in Literatures in modern Indian languages (Delhi, 1951). Much of the controversy regarding the linguistic affiliation of language units such as Maithili, Konkani, Bhili and others could be avoided if the distinction between language units defined by historical and superposed relationships were made clear.

12. Non-linguistic criteria may of course be used to rate languages, e.g. the use of writing, kind of religion, economic structure, etc., but it seems wise not to use the term 'dialect' in this connection. Also, it may well turn out that languages can be rated on certain purely linguistic scales, such as

regularity of morphophonemics, ability to form compounds, or to transfer morphs from one word class to another, etc., but such classification has not yet been systematically carried out and in any case will undoubtedly fail to provide a linear scale of absolute rating for languages as wholes. Cf. R. S. Wells, Archiving and language typology, IJAL 20. 101-7 (1954).

13. For discussions of the rise of standard varieties in the languages of Europe, see J. Vendryes, Language a linguistic introduction to history, tr. P. Radin 260-79 (London, 1931); A. Meillet, Les langues dans l'Europe nouvelle, Chs. VIII-XIV (2nd ed. Paris, 1928). H. Kloss, Entwicklung attempts an analysis of the process of standardization.

14. Other definitions of dialect are, of course, used for special purposes. For example, Bloch in the Postulates defines a dialect for the purpose of that study as a set of idiolects sharing the same phonology.

15. E. Sapir, Language 157-182 (New York, 1921).

16. It must be noted that different kinds of communication very probably have different strengths under this assumption and that this whole area constitutes an important field for sociolinguistic investigation.

17. Cf. L. Bloomfield, Language Chs 3, 19 (New York, 1933).

18. In noting that Grierson did little or no dialect atlas work, we have no intention of belittling his tremendous achievement. The LSI remains one of the world's major productions of linguistic scholarship, and Grierson himself was fully aware of the necessary limitations of his project.

19. Important original work in South Asian dialectology has, however, been done in the detailed historical studies of particular languages or regional dialects. For example, S. K. Chatterji, Origin and development of the Bengali language (Calcutta, 1926); D. Varma, La langue Braj (Paris, 1935); B. Saksena, Evolution of Awadhi (Allahabad, 1938); S. M. Katre, Formation of Konkani (Bombay, 1942).

20. E. R. Leach, Political systems of highland Burma 39 (London, 1954). Cf. also the following statement by Martin Joos, The medieval sibilants, Language 28. 222-31 (1952) quoted in John L. Fischer, Social influences on the choice of a linguistic variant, Word 14. 47-61: "the dialects and idiolects of higher prestige were more advanced in this direction and their speakers carried the drift further along so as to maintain the prestige-marking difference against their pursuers."

21. Cf. also Fischer, op. cit., p. 52, footnote 7. Fischer

states: "Joos seems to me to be unique in his recognition that the two processes (i. e. prestige imitation and the use of linguistic features to emphasize exclusiveness) combine to constitute a self perpetuating cycle."

22. Two examples: R. McDavid, Post-vocalic /-r/ in South Carolina, a social analysis, American Speech 23. 194-203 (1948); S. Sapon, A methodology for the study of socio-economic differentials in linguistic phenomena, SIL 11. 57-68 (1953).

23. Two excellent older studies exist: J. Bloch, Castes et dialectes en tamoul, Mémoires de la Société de Linguistique 16. 1 (1910-11); and L. V. Ramaswamy Aiyar, Tulu prose texts in two dialects, Bulletin of the School of Oriental Studies, 6. 897-931 (1932). One recent study is explicitly concerned with this problem: J. J. Gumperz, Dialect differences and social stratification in a North Indian village, American Anthropologist 60. 668-82 (1958).

24. See G. N. Putnam and E. M. O'Hearn, The status significance of an isolated urban dialect, Language 31, Supplement, Language Dissertation No. 53 (1955) and the bibliography given there.

25. A recent issue of the Indian anthropological journal Man in India was devoted almost entirely to this problem: Man in India 39. 2 (1959).

26. Cf. Bloch, op. cit., Aiyar, op. cit., Gumperz, op. cit. (1958), P. L. Garvin and S. H. Riesenberg, Respect behavior on Ponape, American Anthropologist 54. 201-220. Most linguistic studies of style have been concerned with literary style, but increasing attention is being paid to the broader aspects of style differences in language. For an interesting modern discussion see John L. Fischer, op. cit. Cf. also the symposium on Casual and Non Casual Language held at the American Anthropological Association annual meeting, Chicago, 1957, papers from which were published in International Journal of American Linguistics 24. 253-272; and see now especially T. A. Sebeok, ed., Aspects of Style in Language (New York, 1960). A more traditional but interesting approach to the subject is found in: P. L. Garvin ed., A Prague school reader on esthetics, literary structure and style (Washington, 1955).

27. For example, cf. C. A. Ferguson, Arabic baby talk, For Roman Jakobson 121-8 (The Hague, 1956).

28. For this layer phenomenon in German, cf. W. Schönberger, Die sprachliche Verhältnisse der Tirol-Salzburg-Bayerischen Länderecke, Teuthonista 10. 62 (1934); H. Moser,

Mundart und Hochsprache im neuzeitlichen Deutsch, Der Deutschunterricht 8. 36-61 (1956).

29. J. J. Gumperz, Language problems in the rural development of North India, Journal of Asian Studies 16. 251-9 (1957).

30. Cf. C. A. Ferguson, Diglossia, Word 15. 2 (1959), where this language situation is described in more detail.

LINGUISTIC CHANGE IN SOME INDIAN CASTE DIALECTS

William Bright

University of California

0. Introductory

1. Phonological differences

2. Grammatical differences

3. Lexical differences

4. Conclusions

0. Dialect differences in the languages of India may be thought of in a three-dimensional framework: in addition to the horizontal distribution of geographical dialects, as is found throughout the world, India offers exceptionally clear cases of dialects which are spoken in a single spot, but which may be arranged in a vertical scale correlated with social class. These are the caste dialects, such as those of a North Indian village which have recently been described by Gumperz.[1] In the study of such dialects, the following question may be raised: Is there a correlation between the amount of linguistic change manifested in a dialect and the social status of the people who speak it? In other terms, in what caste dialects are the more archaic features to be found? Investigation of this question may require separate consideration of various parts of the language: that is, phonemic change, grammatical change, and lexical change may not all operate at the same speed in a given dialect.

This paper does not attempt to answer the question completely, but merely to contribute toward an answer, primarily by comparing data recorded from two Kannada dialects of Bangalore District, Mysore State, South India. One is a Brahmin dialect, as spoken by a young woman born and raised in the city of Bangalore. The other is of the agricultural

Okkaliga community, as spoken by a young man from a village a few miles outside the city. Both informants have college educations.[2] The transcriptions are in terms of the following phonemes: /p t ṭ c k b d ḍ j g f s š z m n ṇ ŋ v l ḷ y r h i ī e ē æ ɨ a ā u ū o ō ɔ/.[3] No attempt is made here to describe all existing differences between the two dialects in question; the differences which are described are those best attested in the data.

1. The phonemic systems of the two dialects display the following differences:

1) The Brahmin dialect (hereafter abbreviated as B.) has the phonemes /f z ɔ/ in loanwords from English; in the Okkaliga dialect (O.) these are replaced by /p j ā/ respectively. E.g.: B. kɔfi, O. kāpi 'coffee'; B. ḍazan, O. ḍajan 'dozen.'

2) B. /c/, when intervocalic, sometimes corresponds to O. /s/: B. samācāra, O. samāsāra 'news', from Skt. samācāra-. The other cases noted are also loanwords. The data suggests that O. lacks contrast between /c/ and /s/, at least in medial position. This is reminiscent of the situation in Tamil, where affricate and fricative are members of one phoneme;[4] the situation was probably the same in Primitive Dravidian.[5]

3) B. /š/, occurring mainly in loanwords from Sanskrit, often corresponds to O. /s/: B. šānti, O. sānti 'peace'; B. dēša, O. dēsa, 'country'; B. kašṭa, O. kasṭa 'difficulty.' (The B. forms are identical with the Sanskrit stems in each case.) The lack of contrast between /s/ and /š/ is also a characteristic of Tamil, and presumably of Primitive Dravidian.

4) B. /h/ usually corresponds to O. zero: B. hāku, O. āku 'put'. After pause, B. /he hē ho hō/ correspond to O. /ye yē va vō/ respectively. This is in conformity with morphophonemic changes also operative in B., by which onglides are added to mid vowels after pause, and /vo/ is further replaced by /va/. E.g.: B. hesru, O. yesru 'name'; B. hoge, O. vage 'smoke'; B. hōgu, O. vōgu 'go'. In all these cases, the B. forms represent the earlier stage of development, reflecting PDr. forms with initial *p: cf. OKa. pesar, Ta. peyar, Te. pēru 'name'; OKa. poge, Ta. pukai, Te. poge 'smoke'; OKa. pōgu, Ta. pōku, Te. pōvu 'go'.

5) Aspirated stops of B. correspond to unaspirated ones

in O. The examples are principally loanwords: B. bhūmi, O. būmi 'land' (Skt. bhūmi-).

6) Syllable-initial consonant clusters, occurring in loanwords in B., are rare in O. Thus sequences of sibilant plus oral sonorant in B. correspond to /s/ alone in O.: B. svāmi, O. sāmi 'sir' (Skt. svāmin-); B. manšya, O. mansa 'man' (Skt. manušya-). B. /s/ plus nasal sometimes corresponds to O. /si/ plus nasal: B. snēyta, O. sinēyta 'friend' (Skt. snehita-). B. /kš/ corresponds to O. /c/: B. kšemsu, O. cemsu 'excuse me' (from Skt. √kšam 'to pardon'.)

7) B. /ē/ often corresponds to O. /yā/ when a mid or low vowel occurs in the next syllable: B. pēṭe, O. pyāṭe 'town'. In these cases the pronunciation with /ē/ is older: cf. Ta. pēṭṭai, Marathi peṭh.

8) B. /o/ often corresponds to O. /a/, as in the second morpheme of B. mal-koḷḷ-ōṇa, O. mal-kaḷ-ani 'let's go to bed'. This element, having reflexive meaning in Ka., is cognate with Ta. koḷ 'receive', Te. konu 'take'. Cf. also B. barokke, O. barakke 'for coming'; B. āgolla, O. āgalla 'doesn't become'. Both of these are contracted from forms such as the literary equivalents bar-uvu-da-kke and āg-uvu-d-illa respectively.

9) B. /i̯/ in the sequence /ri̯/, occurring in loans from Sanskrit, usually corresponds to O. /u/: B. sri̯ŋgāra, O. suŋgāra 'beauty' (Skt. śṛŋgāra-.)

The foregoing correspondences may be summarized as follows: Insofar as it fails to accept new phonemic contrasts occurring in loanwords, the O. dialect is the more archaic of the two. It is innovating, however, in its loss of earlier /h/, and in the changes of /ē/ to /yā/ and /o/ to /a/.

2. Grammatical differences between the two dialects include the following:

1) The dative suffix of B. has the allomorphs -kke after /a/ and -Age elsewhere. (A is a morphophoneme commonly actualized as zero after consonants; a vowel plus A, however, becomes immune to the syncope which otherwise affects short vowels.) This corresponds to O. -ke after nouns of all types: B. ūṭa-kke, O. ūṭ-ke (with syncope) 'to dinner'; B. madve-ge, O. maduv-ke 'to a wedding'; B. mārkeṭ-ge, O. mārkeṭ-ke 'to the market'. O. also has the variant -kya: B. pēṭe-ge, O. pyāṭ-kya 'to town'. Historically, B. -kke and -Age may be

traced to OKa. -kke and -ige respectively. OKa. also has the form -ke, however;[6] this has been lost in B., but generalized in O. to the point of replacing other allomorphs.

2) Locative formations in B. contain an element -alli, corresponding to O. -agi: B. baṇḍy-alli (also contracted to baṇḍ̄ili), O. baṇḍy-āgi 'in a cart'. The B. form may be identified with the adverb alli 'there'; the O. form is of uncertain origin, though homonymous with āgi 'having become', from āgu 'become'. Formations in -alli are attested in OKa., along with several other types lacking in the modern dialects;[7] those in -āgi are apparently innovations.

3) The first person plural imperative suffix is B. -ōṇa, O. -āni, -ani, -ana. E.g.: B. hōg-ōṇa, O. vōg-āni 'let's go'; B. mugs-ōṇa, O. mugs-ani 'let's finish'; B. kūt-koḷḷ-ōṇa, O. kuūt-koḷḷ-āna 'let's sit down'. Cf. OKa. -uvam, MKa. -uva, modern written Ka. -uva, -uvana, -uvaṇa, -ōṇa.[8] If, as is claimed by Kittel, /n/ is earlier than /ṇ/ in this suffix, then the O. forms are archaic in retaining it. The vowels of the O. forms, however, appear to be innovations.

4) Certain verb stems show different allomorphy in the two dialects: āgu 'become', but B. āg-ta O. ā-ta 'becoming'; B. hōgu, O. vōgu 'go', but B. hōg-ta, O. vō-ta 'going'; bā 'come', but B. bar-ta, O. bat-ta 'coming'. The O. forms are unattested in older Kannada. A different case, however, is B. kū-tu, O. kun-tu 'having sat' (from a stem occurring only in these forms); both of these date from the medieval period, being attested from 1398 and 1585 respectively.[9]

5) The verb iru 'be' has a set of forms in B. which lack a tense suffix; corresponding to the third person forms of these, O. uses pronouns plus the emphatic morph -e. E.g.: B. iddāne, O. avn-e 'he is' (avnu 'he'); B. iddāre, O. avr-e 'they are' (avru 'they'); B. ide, O. ad-e 'it is' (adu 'it'). All of these are modern formations: of the O. forms, however, only the neuter ad-e is found in the modern written language.[10] The other O. forms may represent an analogical extension on this model.

The foregoing grammatical correspondence thus show the B. dialect to be generally more archaic.

3. Finally, the following cases typify the differences of vocabulary between the two dialects (OKa. and other Dravidian forms are from Kittel's dictionary):

1) In some cases, where there is only a slight phonemic difference between B. and O. forms, the B. form is known to be the earlier one. e.g.:

B. horḍu, O. voldu 'start out'; cf. OKa. poṟa, peṟa 'out', Ta. piṟ 'outer', Malayalam puṟ 'outside', Te. peṟ 'foreign'.

B. bāglu, O. bāklu 'door'; cf. OKa. bāgil.

B. bombe, O. gombe 'doll'; cf. OKa. bombe, Ta. pommai, Te. bomme.

B. matte, O. mante 'again'; cf. OKa. matte, Ta. marṟai.

2) In other cases, the O. form is known to be earlier; there is not necessarily any similarity between the two forms. E.g.:

B. prayāṇa, O. payṇa 'journey'. The B. form is a recent copy of Skt.; the O. form is from OKa. payaṇa, which in turn is from Skt. prayāṇa- or prayaṇa-.

B. bahḷa, O. balu 'much, very'. Cf. MKa. bahaḷa, from Skt. bahaḷa; but OKa. bal 'strength, greatness', Ta. val, Te. balu.

B. togo, O. teko 'take'; cf. OKa. tekkoḷḷ-.

B. hēge, O. yeŋge 'how'; cf. MKa. hēge, but OKa. heŋge.

B. huḷi, O. yesru 'curry'; cf. OKa. esar.

B. ūṭa māḍu, O. unnu 'eat a meal'; cf. OKa. uṇ, Ta. uṇ.

3) In the remaining cases, the relative age of the forms is not clear. E.g.:

B. sumne, O. sumke 'quiet'; cf. OKa. summane and summage, both attested in the same work.

B. yeṣṭu, O. yēsu 'how much', both attested in the same MKa. work.

B. sarti, O. sāri 'time, turn'. Both are attested only in modern times. The B. form may be from Skt. sarat 'going'; with the O. form, cf. Malayalam sāri.

B. solpa, O. rav-asṭu 'a little', both modern. The B. form is from Skt. svalpa; the O. form is literally 'as much as (asṭu) a grain (rave)', cf. Te. rav-aste.

B. snāna māḍu, O. nīr āk-kō 'bathe', both modern. B. is literally 'to make a bath', with Skt. snāna- 'bath'; O. is literally 'to put water on oneself':

B. sinima, O. bayskōpu 'movies', from English cinema and bioscope respectively; the former may be the more recent borrowing.

4. Generalizing from the above comparisons, we may draw the following conclusions: O. accepts foreign words and foreign phonemic patterns less readily than does B. Its vocabulary is thus more archaic, and its phonemic system is likewise archaic in that it lacks certain contrasts which B. has taken over from Sanskrit and other languages. B., on the other hand, is more resistant to phonemic changes of native sounds, and to grammatical change.

Reasons for B.'s receptiveness to importations from other languages are not hard to find: the Brahmins have traditionally been conversant with Sanskrit, and more recently with English. The prestige attached to both these languages has increased the likelihood for bilingual speakers to introduce words from them into Kannada. It is not quite as easy to account for the conservative tendency of B. as regards phonemic and grammatical change. A conditioning factor may be the greater literacy of the Brahmin community; that is, the permanence of the written word may retard change in the spoken language.

In this connection, data from Tulu, another Dravidian language of South India, deserves examination. Although separate dialects of Tulu are spoken by Brahmins and by non-Brahmins or Shudras, neither dialect is commonly written. We may therefore hypothesize that the B. dialect of Tulu will show no greater retardation of phonetic change than the S. (Shudra) dialect. The Tulu data available[11] shows, first of all, some correspondences in which the B. forms are more archaic:

B. initial vowel corresponds to S. zero in a number of cases where PDr. has the initial vowel: B. imboḷu, S. mōḷu 'this woman' (Ta. ival, Ka. ivaḷu); B. eḍaŋgɯ, S. ḍangɯ 'left (hand)' (Ta. iṭai, Ka. yeḍa): B. uṇpu, S. nuppu 'food' (cf. Ta. uṇ, Ka. uṇṇu 'to eat'.)

B. /ṣ/ corresponds to S. /t/ in some cases where PDr. has *c: B. sik-, S. tik- 'be obtained' (Ta. cikku, Ka. sikku); in retaining sibilant quality, B. here appears more archaic.

B. /j/ corresponds to S. /d/ where PDr. has *ñ[12] in B. jeñji, S. deñji 'crab' (Ta. ñaṇṭu, ñeṇṭu, Ka. yēḍi). B. appears more archaic, in that it retains palatal quality.

There are, however, other correspondences in which S. has the more archaic forms:

B. initial /ē/ corresponds to S. /yā/ in some cases where PDr. has *iā:[13] B. ēnuu, S. yānuu 'I' (Ta. yāṉ, OKa. ān); B. ēruu, S. yāruu 'who' (Ta. yār, ār, Ka. yāru.)

B. /s/ corresponds to S. /t/, where PDr. has *t, in B. sōj-, S. tōj- 'appear' (Ta. tōṉru, Ka. tōru.)[14]

In another correspondence, both dialects innovate, and neither can be considered the more archaic: B. /ḷ/ corresponds to S. /r/ where PDr. has *r:[15] B. bālæ, S. bāræ 'plantain' (Ta. vāṟai, OKa. bāṟe, modern bāle); B. kōli, S. kōri 'hen' (Ta. koṟi, OKa. kori, modern kōḷi); B. būḷ-, S. būr- 'fall' (Ta. viṟ-, OKa. būṟ-, modern bīḷu.)

On the basis of the limited data, we may therefore conclude that the difference between B. and S. dialects of Tulu, as regards rate of phonetic change, is less marked than that between the B. and O. dialects of Kannada. A further study of Tulu should be especially rewarding in the investigation of Indian caste dialects.

Notes

1. John J. Gumperz, Dialect differences and social stratification in a North Indian village, American Anthropologist 60. 668-82. (1958)

2. The Brahmin idiolect presented here is that of my chief informant in an intensive study of the Bangalore Brahmin speech, and may be taken as typical. The Okkaliga dialect, on the other hand, was studied only briefly. The data obtained is not 'pure' Okkaliga speech; that is, the informant's idiolect has been influenced by the prestige of the Brahmin dialect. Even so, the salient features of Okkaliga speech show up clearly enough in the data recorded from him.

3. The phonemic system of Kannada is described in detail in my monograph, An outline of colloquial Kannada (Poona, 1958.)

4. Murray Fowler, The segmental phonemes of Sanskritized Tamil, Language 30. 360-67 (1954).

5. T. Burrow, Dravidian Studies VI, Bulletin of the School of Oriental and African Studies (University of London) 12. 132-47 (1947). Primitive Dravidian is hereafter abbreviated PDr.; so also Ka. for Kannada, OKa. for Old Kannada, MKa. for

Medieval Kannada, Ta. for Tamil, Te. for Telugu, Skt. for Sanskrit.

6. A. N. Narasimhia, A grammar of the oldest Kanarese inscriptions 141-46 (Mysore, 1941).

7. Narasimhia, Grammar, 151-54.

8. F. Kittel, A grammar of the Kannada language 148-51 (Mangalore, 1903) The symbol m stands for the Kannada sonne, pronounced (in this position) as /m/ in modern Kannada.

9. F. Kittel, Kannada-English Dictionary 441, 457 (Mangalore, 1894).

10. Kittel, Grammar, 133-34.

11. From L. V. Ramaswami Aiyar, Materials for a sketch of Tulu phonology, Indian linguistics 6. 385-439. The author's phonetic symbols are altered to agree with the transcription used above for Kannada.

12. According to T. Burrow, Dravidian Studies V, BSOAS 11. 595-616 (1946), esp. 605.

13. Burrow, Dravidian Studies V, 599.

14. The complicated history of Tulu initial /t c s d j/ is dealt with by Ramaswami Aiyar, Tulu initial affricates and sibilants, Quarterly journal of the Mythic Society (Bangalore) 22. 259-73 (1932). However, the author does not clearly distinguish the various Tulu dialects represented in his material. Thus, the stem meaning 'to be entangled' is given as cikk-, tikk-, śikk-, and sikk-, only the first being distinguished as 'rare sub-dialectal.'

15. Information on the history of *r̤, and on its modern pronunciation, is given in Bh. Krishnamurti, Proto-Dravidian *z, Indian linguistics 19. 259-93 (= Turner jubilee volume I, 1958.)

TAMIL — LITERARY AND COLLOQUIAL[1]

M. Shanmugam Pillai

Annamalai University

1. Introductory
2. Phonological differences
3. Morphological differences
4. Conclusion

1.0 This is a comparative study of the differences in phonology and in noun inflection between two distinct styles of Tamil, which the present author uses on different social occasions. The literary Tamil (LT) is the one he employs in classroom lectures, in public meetings, and on certain other formal occasions. The vocabulary of this latter style corresponds to his writings in journals. The pronunciation is similar to that of the All India Radio news broadcasts in Tamil and to that ordinarily employed in reading a written text. The colloquial Tamil (CT) is the one he uses in ordinary conversation, when talking to his family and friends and also to students outside the classroom. It is similar to the style employed by other educated native Tamils throughout the present Madras State, but differs from the speech of the uneducated. The dialects used by the latter show a number of regional and caste variations, which have not as yet been systematically studied.[2]

1.1 The LT demands a few more clarifications. Prose writings of the last century and even of the early years of this century were replete with Sanskrit words. During the early decades of this century the purist movement, started by Maraimalai Adigal, which advocated the removal of all foreign words from Tamil, left a permanent impact on modern Tamil prose. Very few writers were left untouched by this impact. Modern Tamil prose including that of the present author, though

not one hundred per cent 'pure', is purer than what it was a few decades ago. These changes however were confined to the lexicon and have had little effect on the structure of the language as such. The LT described here also differs from the rhetorical style used in the speech making of certain political organizations which is characterized by long sentences, complicated syntax and certain characteristic intonation patterns. It is further distinct from the so-called Pandit style replete with archaic technical terms and grammatical features. This latter style is occasionally used by the present author while discussing ancient grammatical treatises but not in talking about more general subjects and writing for the general public.

2. LT and CT have the same inventory of phonemes.[3]

Consonants

	labial	dental	alveolar	retroflex	palatal	velar
Stops	p	t̲	t	ṭ	č	k
	b	d̲		ḍ	ǰ	g
Spirants	f		s		ç	
Nasals	m	n̲	n	ṇ	(ñ)	ŋ
Laterals			l	ḷ		
			r			
				l̲		
Semi-consonants	v				y	

Vowels

i	e	æ	a	o	u
i:	e:		a:	o:	u:

The phonetic correlates[4] of the above symbols are similar for the two styles, with the exception of the word final allophone of /n/, which is [n] in LT and [˜] in CT.

Although the phonemic inventory is the same, the two styles show a number of differences in the distribution of the individual phonemes. These are classified and listed in the section below. Some of the correspondences found are regular, i. e. they are predictable in terms of the LT phonemic

environment. Others are confined to limited lists of items. The historical relationship between the LT and CT forms has not yet been investigated in detail. It was found convenient, however, to give lists of correspondences arranged as they might be arranged in a comparative historical study, with the LT forms taken as primary.[5] Process terminology is occasionally used, although this does not necessarily have historical implications.

2.1 Final consonant correspondences.

CT has no final consonants with the exception of final /n/, (phonically [˜]) and of the final /m/ in /o:m/ 'name of the deity' (a Sanskrit loanword used in prayer). The various correspondences of final LT consonantal themes are listed below.[6]

(1) CV:y regularly corresponds to CV

pa:y	pa:	mattress
te:y	te:	rub
ka:y	ka:	nut
na:y	na:	dog
no:y	no:	disease
va:y	va:	mouth

(2) CVy regularly corresponds to CVyyi

piy	piyyi	tear
ney	neyyi	ghee
čey	čeyyi	do
kay	kayyi	hand
vay	vayyi	put
poy	poyyi	lie (untruth)

(3) Final -ay regularly corresponds to -e

pil̤l̤ay	pil̤l̤e	child
toṇṭay	toṇṭe	throat
čaṭṭay	čaṭṭe	shirt
kaṭṭay	kaṭṭe	block of wood
kaṭay	kaṭe	shop
vaṭay	vaṭe	an edible
malay	male	mountain
naṭay	naṭe	step

(4) Monosyllabics of the structure (C)VC$_1$, where C$_1$ is either /n/, /ṇ/, /l/, or /ḷ/, regularly correspond to (C)VC$_1$C$_1$u.

pen	pennu	pen
peṇ	peṇṇu	woman
kaṇ	kaṇṇu	eye
uṇ	uṇṇu	eat
pal	pallu	tooth
vil	villu	bow
kaḷ	kaḷḷu	toddy
eḷ	eḷḷu	sesame

(5) Final -m corresponds to final -n in some polysyllabic words.

koṭam	koṭon	pot, vessel
paṭam	paṭon	picture
ṯi:rkkam	ṯi:rkkon	valor

(6) In the following items (C)Vm corresponds to (C)Vmmo.

nam	nammo	our
um	ommo	your Cf. (19)

(7) In one other item V:m corresponds to CV:mo.

na:m	ma:mo	we

(8) In most other themes the final -C$_1$ corresponds to -C$_1$u. For exceptions, cf. 3.2, 3.4.

pa:l	pa:lu	milk
ṯo:l	ṯo:lu	skin
a:ḷ	a:ḷu	person
ma:n	ma:nu	deer
aṇil	aṇilu	squirrel
kaṭal	kaṭalu	sea
pensil	pensilu	pencil
u:ḻ	u:ḻu	fate
uyir	uyiru	life
ṯayir	ṯayiru	yogurt
avar	avaru	he (remote)
ivar	ivaru	he (proximate)
kaṭavuḷ	kaṭavḷu	god

2.2 Non-final consonant correspondences.

(9) CT has no intervocalic -tt-. LT medial -tt- regularly corresponds to -tt.

a:ttil	a:ttle[7]	in the river
ka:ttil	ka:ttle	in the air
u:ttu	u:ttu	spring, fountain
ma:ttu	ma:ttu	change
e:ttu	e:ttu	drive
uttu	uttu	(looking with) keenness
vitta:n	vitta:n	he sold
vatte	vatte	to dry

(10) Medial -tC_1- corresponds to -C_1C_1-.

kaṭci	kacci	party, association
paṭci	pacci	bird
ke:ṭke	ke:kke	to hear
uṭka:rrnta:n	okka:nta:n	he sat Cf. also (15, 19)

(11) Medial -kk- after LT -y- or -i- regularly corresponds to -kky-.

asaykke	asekkye	to shake
aṭikke	aṭikkye	to beat
paṭikke	paṭikkye	to learn
paṭaykke	paṭekkye	to offer to gods

(12) Intervocalic ṉnV, llV, l̤lV, and vvV regularly correspond to ṉ, l, l̤, and v before consonants.

kaṉṉil	kaṉle	in the eye (locative)
kollu̱va:n	kolva:n	he will kill
kol̤l̤uva:n	kol̤va:n	he will have
evvaḷavu	evḷavu	how much

(13) Medial -l̤- before nt regularly corresponds to zero.

avul̤ntu	avuntu	untied
kavul̤ntu	kavuntu	toppled

(14) Medial y after -a- in polysyllabic words regularly corresponds to zero; the -a- appears as -e- Cf. also (3).

aṭaykke	aṭekkye	to shut

uṭaykke	oṭekkye	to break
paṭayṯṯa:n	paṭečča:n[8]	he offered (to god)

(15) Medial -rr- corresponds to zero before C_1C_1 or nasal plus homorganic consonant in the items below.

utka:rrnta:n	okka:nta:n	he sat
nirrkkirra:n	nikkrra:n	he is standing
pa:rrkkirra:n	pa:kkrra:n	he is seeing
pa:rrṯṯa:n	pa:ṯṯa:n	he saw

(16) In a few other items medial -rr- corresponds to -l- before geminates.

parrppoti	palppoṭi	tooth powder
karrttu:ṇ	kalttu:ṇ	stone pillar
karrččaṭṭi	kalččaṭṭi	stone vessel

(17) C_1VC_2 corresponds to C_1C_2 in some items.

kaṇṇil	kaṇṇe	in the eye Cf. also 3.4
o:ṭina:n	o:ṭṇa:n	he ran
ca:ṯṯina:n	ča:ṯṯna:n	he pointed to him (for something)

(18) Initial m- corresponds to v- in the following word.

mi:say	vi:se	mustache

2.3 Vowel correspondences.

(19) /i/ and /u/ in the first syllable of polysyllabic words correspond to /e/ and /o/ respectively before -a- in the next syllable. The -a- before final LT -m becomes -o- before final -n; cf. (27). Elsewhere it appears as /e/.

iṭam	eṭon	place
iṇaŋku	eṇeŋku	agree
tirrappu	terreppu	opening
pirrappu	perreppu	birth
piraṇṭay	pereṇṭe	a creeper
ulakam	olekon	world
urraŋku	orreŋku	sleep
kuraŋku	koreŋku	monkey
pulayppu	poleppu	a living Cf. also (14).
vilay	vele	price Cf. also (3).

kiḷay	kele	branch
kulay	kole	killing
kuṭay	koṭe	umbrella

(20) Initial LT i- corresponds to CT zero in the items below.

impa	mpa	a call for cows
inta	nta	expression used when giving something to someone
iraṇṭu	reṇṭu	two Cf. also (19).

(21) In one other instance e: corresponds to ya:

e:n	ya:n	why

(22) The above correspondences are not found in the word ivar 'he', and before geminate consonants.

ivar	ivaru	he
ivarukku	ivarukku	to him
kupparre	kupparre	upside down
kuṭṭacci	kuṭṭačči	short girl
kuṭṭay	kuṭṭe	small pond
tillay	tille	name of a place
pillay	pille	child
killay	kille	parrot

(23) Medial -i- is lost in CT after l or rr plus geminate tt. Cf. also (9).

va:lttina:n	va:lttna:n	he praised
ta:lttina:n	ta:lttna:n	he lowered
nakarrttina:n	nakattna:n nakarrttna:n	he moved (trans.)
patarrttina:n	patattna:n patarrtna:n	he spread (trans.)

(24) Medial u is lost in these two items.

tamilukku	tamilkku	to or for Tamil
kaṭavuḷ	kaṭavḷu	god

(25) C_1iC_2- correspond to C_1C_2- in the following words.

pila:vu	pla:vu	jack tree

miḷa	mḷa:vu	an animal
kiṭa:vu	kṭa:vu	buffalo
kiḷappu	kḷappu	hotel

(26) In many polysyllabic words the vowel in the second syllable corresponds to zero.

pa:kkiyam	pa:kkyon	personal name
o:ṭina:n	o:ṭna:n	he ran
pa:ṭina:n	pa:ṭṇa:n	he sang
evvaḷavu	evḷavu	how much
kolluva:n	kolva:n	he will have
kolluva:n	kolva:n	he will kill
kaṇṇil	kaṇḷe	in the eye
ma:ppiḷḷay	ma:ppḷe	son-in-law
kaṯṯina:n	kaṯṯna:n	he cried

(27) /-a-/ before final -m ordinarily corresponds to /o/ before final -n. This vowel change is connected with final consonant correspondence (5).

koṭam	koṭon	pot
paṭam	paṭon	picture
ṯi:rkkam	ṯi:rkkon	valor

(28) Final -an corresponds to -en (phonetically [ɛ̃]) in the items below.

avan	aven	he
čaṉṯiran	čaṉṯren	moon
ivan	iven	he
makan	maken	son

(29) -an corresponds to -on in the following single item.

kaṭan	kaṭon	debt

(30) Final -u corresponds to final -e in these two items.

iŋku	iŋke	here
eŋku	eŋke	where

2.4 Some correspondences involve sequences of consonants and vowels. Correspondences (3), (14) and (5), (27) might have been listed in this category. Others are:

(31) Medial ak corresponds to -o-.

marumakan	marumoen	son-in-law
marumakaḷ	marumoa	daughter-in-law

(32) In the following words final -lutu corresponds to zero or final n ([˜] phonetically).

ippoḷḷutu	ippo˜ippon	now
appoḷḷutu	appo˜appon	then
eppoḷḷutu	eppo˜eppon	when

(33) In two other instances CT has final -vu where LT has zero.

miḷa:	mḷavu:	an animal
kiṭa:	kṭa:vu	buffalo

3. Tamil nouns are inflected for gender, number and case in that order. The arrangement of the suffixes in an inflected noun may be represented by A B C D E, where the symbols have the following values:

A stem
B stem formative if any
C gender suffixes if any, neuter is unmarked
D number, singular is unmarked
E case suffixes if any

Morphemes B-E are all suffixal morphemes occurring in the same relative order in LT and in CT. The two systems show differences, however, both in the shape of individual morphemes and in morpheme distribution. Some of these are illustrated in the section below.

3.1 Stem formatives.[9] Certain classes of nouns have morphs between the stems and the gender suffixes to which no meaning can be ascribed. These we call stem formatives; in Hockett's terminology they are empty morphs.[10] LT has -tt̲-, -čč-, and -ṭṭ- as stem formatives. All of them occur in CT. However, the distribution in CT differs from that in LT.

In LT, nouns in the neuter class which end in -am have the stem formative -tt̲- in the singular before case suffixes and this stem formative does not occur in the plural with or without case suffixes. But in CT this stem formative is found both in the

singular and in plural before case suffixes. In the plural with case suffixes the alternant form without the stem formative is also used. When there is no case suffix there is no stem formative either in the singular or in the plural.

	Stem	sg. with case suff.[11]	pl. without case suff.	pl. with case suff.	
LT	maram	marattay	maraŋkal̤	maraŋkal̤ay	tree
CT	maron	maratte	maraŋko	marattuol̤e or maraŋkal̤e	
LT	paṭam	paṭatty	paṭaŋkal̤	paṭaŋkal̤ay	picture
CT	paṭon	paṭatte	paṭaŋko	paṭattuol̤e or paṭaŋkal̤e	

The other stem formatives -čč- and -ṯṯ- have similar distributions in LT and CT.

3.2 Gender. Neuter is unmarked both in LT and CT. The masculine gender suffixes in LT are an, a:n, and ∅ (zero). Except for the phonologically conditioned morphophonemic alterations there is no difference whether they occur finally or non-finally.[12] In CT the suffix -an has the allomorph -en in final position, -an- in non-final position. Cf. correspondence (28). Thus, LT forms to:l̤an, to:l̤anay 'friend (male)' correspond to CT forms to:l̤en to:l̤ane.

The same type of difference as in the masculine is found in the distribution of the feminine suffixes. The feminine suffixes in LT are al̤, a:l̤, i, and ∅, and they are the same whether they occur finally or non-finally. Allomorphs of the suffixes al̤ and a:l̤ in CT are conditioned by position: the form in final position drops the -l̤. Thus, LT forms makal̤, makal̤ay 'daughter' correspond to CT maka, makal̤e; and LT akka:l̤, akka:l̤ay 'elder sister' correspond to CT akka:, akka:l̤e.

3.3 Number. Singular is unmarked both in LT and CT. The noun plurals in LT may be classified into four classes according to the suffixes they have. The plural suffixes in LT are kal̤, kkal̤, ar ~arkal̤, and a:r ~a:rkal̤.[13] There is no difference whether these suffixes occur finally or non-finally.

In CT the nouns fall under eight different classifications. The plural suffixes are different whether they occur finally or non-finally. The two systems are tabulated below illustrating the difference. Numbers I, II, III, and IV in the following classification indicate the classification of the nouns in LT and the subdivisions a, b, c within the numbers indicate the classification in CT of that particular class in LT.

		Final	Non-final	
Ia	LT	-kaḷ	-kaḷ-	
		maraŋkaḷ	maraŋkaḷay	tree
	CT	-ko	uoḷ ~kaḷ ~kḷ	
		maraŋko	marrattuole ~ maraŋkaḷe ~ maraŋkḷe	
Ib	LT	-kaḷ	-kaḷ-	
		makkaḷ	makkaḷay	child
	CT	-ko	kaḷ ~ kḷ	
		makko	makkaḷe ~ makkḷe	
Ic	LT	-kaḷ	-kaḷ	
		ma:ṇavikaḷ	ma:ṇavikaḷay	student (fem.)
	CT	-yo	yoḷ	
		ma:ṇoviyo	ma:ṇaviyoḷe	
IIa	LT	-kkaḷ	-kkaḷ	
		pu:kkaḷ	pu:kkaḷay	flower
	CT	-kko	-oḷ- ~ -kkaḷ- ~ -kkḷ-	
		pu:kko	pu:oḷe ~ pu:kkaḷe ~ pu:kkḷe	
IIIa	LT	-ar ~ -arkaḷ	-ar- ~ -arkaḷ-	
		avar ~ avarkaḷ	avar ~ avarkaḷ	they (no gender)
	CT	aŋko ~ aŋke	-kaḷ- ~ -kḷ-	
		avaŋko ~ avaŋke	avaŋkaḷe ~ avaŋkḷe	
IIIb	LT	-ar ~ arkaḷ	-ar- ~ -arkaḷ-	
		avar ~ avarkaḷ	avar ~ avarkaḷ	they (masc.)
	CT	-uo	-uoḷ-	
		avanuo ~ avaḷuo[14]	avanuoḷe ~ avaluoḷe	they (fem.)
IIIc	LT	-ar ~ -arkaḷ	-ar ~ -arkaḷ-	
		ma:ṇavar	ma:ṇavaray	student (male)
		ma:ṇavarkaḷ	ma:ṇavarkaḷay	
	CT	-uo ~ ko ~ ke	uoḷ ~ kaḷ ~ kḷ	
		ma:ṇavanuo	ma:ṇavanuoḷe	

	Final	Non-final	
	ma:ṇavaŋko	ma:ṇavaŋkaḷe	
	ma:ṇavaŋke	ma:ṇavaŋkḷe	
IVa LT	-a:r ~ -a:rkaḷ	-a:r- ~ -a:rkaḷ	
	na:ṭa:r ~ na:ṭa:rkaḷ	na:ṭa:ray ~ na:ṭa:rkaḷay	name of a caste
CT	-uo ~ ko ~ ke	-uoḷ- ~ -kaḷ- ~ -kḷ-	
	na:ṭa:nuo	na:ṭa:nuoḷe	
	na:ṭã:ŋko	na:ṭa:ŋkaḷe	
	na:ṭã:ŋke	na:ṭa:ŋkḷe	

Except for the free variations in III and IV the plural suffixes in LT are always kaḷ or kkaḷ. In CT, however, the non-final allomorphs of the plural suffixes always end in ḷ, and those of final position end in either -o or -e.

The demonstrative pronouns avan 'he' (remote), ivan 'he' (proximate), the interrogative pronoun evan 'who', and the noun makan 'son', have three forms in LT: masculine sg., feminine sg., and plural without gender. The plural may refer to masculine plural, feminine plural, or both. In CT, however, they have five forms: masculine sg., feminine sg., plural without gender, masculine plural, and feminine plural. They are given below:

LT

masc. sg.	fem. sg.	pl. no gender
avan	avaḷ	avar ~ avarkaḷ
ivan	ivaḷ	ivar ~ ivarkaḷ
evan	evaḷ	evar ~ evarkaḷ
makan	makaḷ	makkaḷ

CT

masc. sg.	fem. sg.	pl. no gender	masc. pl.	fem. pl.
aven	ava	avaŋko	avanuo	avaḷuo
iven	iva	ivaŋko	ivanuo	ivaḷuo
even	eva	evaŋko	evanuo	evaḷuo
maken	maka	makko	makanuo	makaḷuo

3.4 Case morphemes. Differences in the case suffixes between the two systems are confined to the shape of cognate morphemes. The correspondence in the accusative case suffix (-ay corresponding to -e) has been covered under the general phonological changes (3). The rest of the case suffixes, which do not come under the phonological correspondences are tabulated below in the two systems.

	LT	CT	
Inst.	-a:l	-a:le	
	maraṯṯa:l	maraṯṯa:le	by the tree
Conj.	-o:ṭu	-o:ṭe	
	avano:ṭu	avano:ṭe	with him
Dat.	-kku	-kki	after front vowels /i/ or /e/ in the preceding syllable; -kku elsewhere
	panaykku	panekki	to the panay tree
	maṇikku	maṇikki	to Mr. Maṇi
	avanukku	avanukku	to him
Abl.	-iliruṉṯu	-leruṉṯu	
	maraṯṯiliruṉṯu	maraṯṯleruṉṯu	from the tree
Gen.	-uṭaye	-o:ṭe	
	ennuṭaye	enno:ṭe	my
Loc.	-il	-le	
	maraṯṯil	maraṯṯle	in the tree
'among'[15]	-il, -ukkuḻḻe	-ḻe	
	maraŋkaḷil	maraŋkaḻḻe	
	maraŋkaḷukkuḻḻe	maraṯṯuoḻḻe	

Similar differences both in the shapes and in the distribution of the morphemes are found in the verbal system also. čeyṯa:y 'you did' and piriṉṯa:y 'you separated (intransitive)' which have the same past tense suffix -ṯ- come under a single class in LT. But in CT they come under two different classes since the first one has the past tense -ṯ- whereas the second one has the past tense suffix -c-: čeyṯe 'you did', piriñče 'you separated'. There is, in fact, no past tense -č- suffix in LT.

4. Conclusion

The above data, incomplete as they are, should be sufficient to illustrate the distinct nature of the two styles. From

the point of view of history they seem to represent two different stages of development of the language existing side by side. A complete explanation of the historical relationship between the two will require more comparative study and above all additional information on local dialects which have not been studied so far. The results of such a study, aside from being of interest to historical linguists, should also provide valuable information regarding population movements and social change.

The existence of the two styles also has a number of more practical implications. Aside from the features illustrated here there are of course a number of additional differences in vocabulary and syntax. Acquisition of literacy in Tamil therefore requires much more than learning the alphabet. It is more like learning a second language. This fact has not been considered so far in the teaching of literacy and the language arts. Teaching materials which take account of the realities of spoken and literary styles could effect great improvements in the efficiency of present methods.

Notes

1. I would like to express my gratitude to The Rockefeller Foundation for a fellowship which enabled me to study linguistics at Cornell University, Ithaca (1956-57) and at the University of California, Berkeley (1957-58) and particularly to Dr. John J. Gumperz, University of California, Berkeley, who suggested and helped in the preparation of this paper, and Dr. Murray B. Emeneau, who read the first draft.

2. Sir George A. Grierson's Linguistic Survey of India, Calcutta, 1927, did not cover the Tamil speaking areas of Madras State and Travencore Cochin. Some features of Tamil caste dialects are described in Jules Bloch, Castes et dialectes en tamoul, Mémoires de la Société de Linguistique 16. 1-30 (1909). In regard to the uniformity of speech among educated Tamils, note Bloch's statement that special features of Brahmin speech are dropped in contacts with non-Brahmins (ibid. 29).

3. There are two recent publications dealing with the phonemics of Tamil; Murray Fowler, The segmental phonemes of Sanskritized Tamil, Language 30. 362-7 (1954), and Gordon H. Fairbanks, Frequency and phonemics, Indian Linguistics 17. 105-13 (1957). Fowler's analysis (from here on referred to as ST) seems to represent a more formal variety of Tamil,

closer to the Pandit style. His system consists of five unitary vowels /i e a o u/, occurring as simple vowels or geminates and 22 consonants /p b t d ṭ ḍ k g m n ṇ l ḷ v c r̲ x h w y r l̲/. Fairbanks' system represents the CT as given here. In the footnotes to his paper Fairbanks has pointed out some of the differences between his and Fowler's analysis. We will here list those differences not mentioned by Fairbanks.

ST does not have /æ/. In LT and CT /æ/ is found in words borrowed from English: /bænk/ 'bank', /bæt/ 'bat'. /h/ in ST corresponds to [x~ɣ] in LT and CT, an allophone of /k/.

The contrast between /x/ and /c/ in ST [k'ʌxcɨ] 'a measure of weight', and [k'ʌttʃɨ] 'girdle' is not found in CT for the simple reason that /kaxču/ is a literary word, not now in use. In LT the above are [k'ʌhsɨ] and [k'ʌčč'ɨ] where [h] may be analyzed as an allophone of /k/ occurring before fricatives, the allophones of the respective stops and before retroflex flap [ṛ] allophone of /t/. Other citations illustrating this allophone are [ʔʌhðɨ] 'that', [ʔʌhxi] 'shrink' etc. All these forms including [k'ʌhsɨ] are found in ancient literature and not used in modern Tamil.

The phonemes /f/ and /ç/ are found only in borrowed words, /f/ in words like /film/ 'film', /ka:fi/ 'coffee', /so:fa/ 'sofa', etc. all borrowed from English and /ç/ only in the word /çinti/ 'Hindi'. In the speech of the uneducated /çinti/ is /inti/.

The ST continuant /r/ is analyzed as the geminate of the flap /ṛ/.

ST has only three nasals /m n ṇ/. Both in LT and CT six nasals have to be recognized, (or five if a different solution is adopted). All the six nasals occur initially when referring to the letters in the Tamil alphabetic system.

LT	CT	
makaram	ma:na:	the letter m
n̲akaram	n̲a:na:	the letter n̲
nakaram	na:na:	the letter n
ṇakaram	ṇa:na:	the letter ṇ
ñakaram	ña:na:	the letter ñ
ŋakaram	ŋa:na:	the letter ŋ

Children are taught to distinguish these sounds in school. Of these, /m n ṇ/ contrast in many environments. /ñ/ occurs only in the above example and in ña:yirru 'sun, Sunday', contrasting with nya:yam 'justice'; it might also be analyzed as /ni/, thus reducing the number of nasal phonemes to five. /n/ contrasts with n before /k/ e. g. ponkay 'golden hand' t̲aŋkay 'younger sister'. n̲ is found only in the name of the letter quoted above.

4. Phonetic values of the above symbols are given in

Fairbanks (ibid). The Fairbanks analysis has seven nasal vowels /ɛ̃ ɛ̃: ã: õ õ: ũ ũ:/ occurring finally. These are analyzed as -Vn here. This is possible since we set up the two styles as two distinct systems. However, if we were to present an overall analysis of Tamil, we should have to set up the nasal vowels as separate phonemes.

5. In all the statements and in the examples that follow the LT forms are given first and then the CT forms.

6. The statements of phonological correspondences are numbered consecutively in parentheses (1) to (33) through the subsections 2.1 - 2.4.

7. -il and -le are the LT and CT locative case suffixes (cf. 3.4).

8. -tt- and -čč- are the LT and CT past tense suffixes.

9. The term 'stem formative' is an extension of the term 'formative suffixes' used by M. B. Emeneau, Kolami - A Dravidian Language (Berkeley, 1955). These suffixes could be analyzed as part of the stems or as part of the suffixes that follow. But the existence of classes of nouns without these suffixes suggests the segmentation of these morphs though no meaning can be ascribed to them.

10. C. F. Hockett, Problems of morphemic analysis, Language 23. 321-43 (1947).

11. The examples are given only with the accusative case suffixes. The situation is the same for all the cases.

12. Before velar consonant /n/ has the morphophonemic alternant /ŋ/.

13. In old Tamil kaḷ is the plural suffix in neuter nouns and ar, a:r are the plural suffixes in non-neuter nouns. In LT both forms ar~arkal and a:r~a:rkaḷ are used though traditional Pandits would still insist upon the use of only ar and a:r.

14. The -an in avanuo is the masculine suffix; -aḷ- in avaḷuo is feminine suffix; aŋ in avaŋko ~ avaŋke should be historically the masculine suffix. But when the form avanuo was used for the masculine plural the form avaŋko ~ avaŋke lost its reference to masculine gender and is now used for the plural without gender, either masculine, or feminine or both. Hence aŋko ~ aŋke is taken to be the plural suffix.

15. The word 'among' is not a case name. It gives the meaning of this case form, which occurs in the plural.

LITERARY AND COLLOQUIAL BENGALI IN MODERN BENGALI PROSE

Edward C. Dimock

University of Chicago

1. Introductory
2. Linguistic differences
3. History of the controversy

1. It has long been recognized that there are at least two distinct forms of the Bengali language, the sādhu - bhāṣā (literary language, hereafter SB) and the calit - bhāṣā (the colloquial language, hereafter CB). This condition of having two different forms of a language, with the distinction between them relating to prestige or other extra-linguistic values, is not an unfamiliar one. Bloomfield mentions it in his description of language community.[1] Historically and within our own culture, we need look no farther than to the relationship between Latin and the vernaculars in medieval Europe.[2] And, on a somewhat different level within our own contemporary society, there is the prestige-giving value of certain types of usages, pronunciations, and "accents", among certain groups of people. With this we are all familiar and I am sure that examples need not be multiplied.

There is, however, one aspect of this general type of linguistic situation which might not be familiar to English-speaking people: that for a given language there might be one form used almost exclusively for literature and another form used almost exclusively on the conversational, social-intercourse level. Yet this is a feature of a large number of the languages of the world, including Bengali. A situation of this type in Arabic, even more radical than that in Bengali, has been recently described by Al-Toma.[3] He establishes certain categories for the definition of the distinction between two forms of Arabic, and these categories might be adopted by us here, in order to

open our discussion of the similar distinction in Bengali.

In the first place, Al-Toma states that Classical Arabic is the language of formal communication, and that its usage is determined by "situations which have the character of formality." His second generalization is that Classical Arabic is uniform throughout the Arabic-speaking world. And, thirdly, he says that Classical Arabic has more prestige than does the colloquial. The Bengali SB also has these characteristics; but to make the statement entirely accurate for Bengali, a few modifications would have to be made. In the first place, it must be understood that pure SB, unlike Classical Arabic, is rarely spoken.[4] Al-Toma mentions as "situations with the character of formality" lectures, speeches, formal addresses, and the like. In Bengali, what would be used in these types of situations would be formal or perhaps SB vocabulary with predominantly colloquial pronunciation and grammar. This almost exclusively written character of the SB should be borne in mind when thinking of the universal uniformity of the SB throughout the Bengali-speaking world. In the second place, although it is possible in spoken Arabic to have a mixture of Classical and colloquial, in written Arabic Classical and colloquial are never mixed. As we shall see, the situation is quite different for written Bengali.

It might be put this way: Classical and colloquial Arabic may be considered in some sense as two different languages, with clear and linguistically definable differences between them. The distinction between SB and CB is not as clear-cut. SB and CB might be better thought of as opposite poles of the same language.[5]

Pure SB is the recognized norm for the writing of the language, the peculiarities of which are respected and understood wherever Bengali is written or read. The CB, on the other hand, is first and foremost the standard spoken language of Calcutta. It is not merely another regional dialect.[6] It is by no means limited to Calcutta or to the area immediately surrounding that city, but is used and understood throughout an area which crosses over many dialect boundaries. In relatively recent times there has been a movement within the Bengali language toward the fusion of the two forms. In modern Arabic, I am told, only the most daring of dramatists allow their characters to speak in the colloquial form, even when the characters

might be of a category in which the speaking of Classical Arabic would be absurd from a realistic point of view. This situation prevailed in Bengali literature about half a century ago.[7] It does no longer. Although there had been writers before him who had used the CB for serious literature,[8] it can perhaps be fairly said that Rabindranath Tagore began a trend which continues strongly in the present day.

Tagore is fond of illustrating social and cultural situations with little fables. He tells such a fable to point out the increase of the use of CB in literature.[9] The story is that of a king who had two queens. One was sophisticated, immaculately groomed, and utterly regal. The other, of course, was simple, straightforward, and charming in her unpretentiousness. I am sure that everyone has read enough fairy-tales to be able to imagine the course of the story. At first, the younger, un-queenly wife was scorned, but gradually she began to charm the king and to win his favor. And Rabindranath says, "The end of the story has not yet come; but there will be a day when the younger will sit upon the throne alone." There might be some who would question the conclusion, but I am quite sure that the analogy is true enough.

2. If one considers that the SB and the CB are at the two opposite poles of the same language, it is clear that descriptively there is an infinite number of points between them. Many of these points are occupied by individual writers whose work represents the state of the language at some phase in the history of the movement of the two poles toward each other. If one could graph the accepted norms of the SB and CB over the last century, I suspect the graph would show a truncated pyramid. Starting from the two extremes, the lines would show a gradual merging. The graph, however, is incomplete. The apex of the pyramid, the identity of the SB and CB, lies somewhere in the future. It is clear that over the last century the two extremes have been moving more and more toward the center. At the present day, they are still definably distinct, but are considerably closer together than they were a century ago.

Tagore points out that in some ways the CB is replacing the SB, but that at the same time the CB is taking on some of the characteristics of the SB.[10] One of the examples he uses is this: some years ago, a man of the servant class using the

term /ɔpekkha/[10a] ("expectation of or waiting for someone or something") would have been considered laughably absurd. This, Tagore says, is no longer so. Although /ɔpekkha/ is a purely SB term, it is becoming entirely acceptable for all people.[11]

It is clear that if this study were to be a complete one, it would not only have to be descriptive of the present state of the two forms of the language, but it would have to trace the movement of the two forms toward juncture over a period of time. Account would also have to be taken of the fact that from the point of view of modern Bengali writers, the controversy between SB and CB is still raging. Although in some speech and in most writing CB and SB have come close together, there are still ardent and powerful apologists for the extremes of both. Buddhadeva Bose, for example, a most influential writer and critic in modern Bengal, feels very strongly that modern creative literature should be written entirely in the spoken language, with care taken to rid it of any influence of the SB. As we shall see, there are apologists for the other extreme, as well. Tagore represents the middle ground, which is probably also the realistic ground, that there is no need for concern over the blending of CB and SB. He feels that their blending is a natural linguistic development, and what once was a literary crime of major proportions, the mixture of CB with SB, can no longer be considered such.[12]

Descriptively, the differences between the SB and CB forms of the language can be summarized under four general categories: lexical, phonological, morphological, and syntactic.

2.1 Lexical distinctions. On the basis of a word list devised by Professor Gordon H. Fairbanks,[13] it was found that, of 511 items counted, 39.5% have correspondence (i.e., that the item is identical for CB and SB), and 5.4% can be used optionally in either CB or SB (i.e., for a given English gloss there are two Bengali meanings, one differing from CB to SB, the other corresponding). Therefore, 55.1% of the lexical items counted always have differing forms in CB and SB. It is, however, probable that these figures are descriptively less than accurate. When the forms were elicited, I had in mind a definition of the breadth of the distinction between CB and SB rather than description of the actual current situation. In all likelihood, these figures represent the point at which CB and

SB are farthest apart, the point which might have been descriptively accurate 75 or 100 years ago. Since I was aiming at contrasting forms, it is more than possible that my informant has given obscure SB forms, forms which are possible but used only rarely, in places where in present-day writing a CB form might more ordinarily be used. Thus the category of optional forms would probably be far greater than the present figures indicate; indeed, this would be reflected by Tagore's statement that the actual lexical differences between SB and CB are diminishing. It might even be said that to a large extent in modern writing a choice between a CB and an SB form is stylistic rather than categorical, and would depend on whether the subject matter was heavy or light and on the author's choice of treatment of it.[14]

A few examples of CB-SB lexical differences are the following:

SB form	CB form	gloss
hɔŋšo	hãš	goose
brikkho	gač	tree
putro	čhele	son
našika	nak	nose
tɔpto	gɔrom	hot
dɔŋšɔn kɔra	kaṃṛano	(to) bite
agɔmɔn kɔra	aša	(to) come
dan kɔra	deoa	(to) give
briṣṭi hɔoa	jɔl hɔoa[15]	(to) rain
ɔpekkha kɔra	daṛano[16]	(to) wait for

It is to be noted that certain of the items above, such as the first and fourth, seem to show a relationship between SB and CB forms. About 18.5% of the total number of forms show such a relationship. The historical relationship between, for example, /hɔŋšo/ and /hãš/ is that /hãš/ represents a normal development of the Sanskrit haṃsa through the Prakrits, while the SB form /hɔŋšo/ represents a re-borrowing by Bengali from the Sanskrit. This is by no means an unusual linguistic phenomenon; grammarians call the direct, popular descendant form tadbhāva and the learned borrowing tatsama.[17] The large number of CB-SB correspondences of this type in Bengali could, I think, be described in a very few statements.

2.2 Phonological distinctions. As the SB is not a spoken language, phonological differences between it and the CB are not significant. Even so, it is interesting to notice that there are people who will pronounce SB differently from their ordinary speech when reading an SB text aloud. The most frequent and noticeable phonetic differences are in terms of consonant clusters. For example:

SB form	CB form	gloss
lɔksmi	lokkhi	Laksmi
smɔrɔn	šɜ̃rɔn	memory

If a general statement were to be made, it might be said that the SB phonologically is closer to the orthography than is the CB. For example, as S. K. Chatterji indicates in the introduction to his Bengali Reader, what is sometimes called the "short a" vowel (phonetically), is, under certain circumstances in the CB, phonetically [o]. In SB pronunciation this vowel tends to remain [ɔ]. There are many other cases of this kind. Orthographically there are three sibilants, dental, palatal and retroflex, which some people attempt to differentiate when reading SB. Phonemically in the CB there is only one sibilant. Further, there are symbols for both dental and retroflex nasals, which some people attempt to differentiate; phonemically there is only one dental. There are several symbols for the velar nasal. The -Cy- cluster is phonemically -CC-, etc. This is an area of the study which to my knowledge has not yet been adequately described.

2.3 Morphological distinctions. Although lexical items represent the major difference between the SB and the CB, significant differences occur on the morphological level. These morphological differences are primarily in verb forms and in certain noun endings. Both of these types of morphological differences can be demonstrated by a translation of the simple expression "He has given (it) to us";

SB /amadigɔke dan kɔriyache/ : CB /amader diyeche/

These differences are firm and consistent throughout the language. A SB verb form is never used in a CB environment, and a SB substantive suffix is rarely if ever used in a CB

environment. The distinctions can be summarized simply:

a. Verb forms (Synopsis of 1st person forms for all tenses):

Tense	SB form	CB form
Simple present	čɔli	čoli
Simple past	čɔlilam	čollam
Conditional	čɔlitam	čoltam
Simple future	čɔlibo	čolbo
Pres. continuative	čɔlitechi	čolchi
Past continuative	čɔlitechilam	čolchilam
Present perfect	čɔliyachi	čolechi
Past perfect	čɔliyachilam	čolechilam
Infinitive	čɔlite	čolte
Conjunctive	čɔliya	čole

b. Noun endings. The most striking difference is the form of the animate plural suffix -ra when followed by a case ending: in SB it is -dig(ɔ)- and in CB it is -de-. Thus, using the first person plural pronoun as an example:

Case	SB form	CB form
Nominative	amra	amra
Genitive	amadiger	amader
Objective	amadigɔke	amader, amadike, or amaderke

There are differences, also, in the definitives and pluralizing suffixes of nouns.

There are other morphological distinctions. Most distinctions, however, have either disappeared or are disappearing in the blending of SB and CB with each other. For example, people writing in SB today pay little if any attention to such considerations as gender, as the pandits once did. This is also true of compound words. SB was once full of compounds, but it is now usual to eliminate sandhi, break down compounds into their component parts, and to treat these component parts as separate entities.[18] It is also generally true that such prefixes as /su-/, /ku-/, /ɔ-/, which indicate quality or negativeness, are rarely found in CB constructions, being considered characteristic of SB.

2.4 Syntactic distinctions. Since to my knowledge not much work has yet been done on Bengali syntax, it is a little difficult to make a systematic statement about it here. A few generalizations, however, might be of some value.

It is possible to consider such verbal compounds as /dan kɔra/ and /ɔpekkha kɔra/ as syntactic phenomena. Such compounds occur in far greater number in SB than in CB. Of the 142 verbs in the list mentioned above, 57 forms were compound in both SB and CB; and 13 were simple in both SB and CB. In no instance did CB show a compound verb where the SB verb was simple.

A second generalization might be that word order in SB constructions is more rigid than it is in CB.[19] Almost invariably in SB constructions, the verb form is final in the sentence. In CB, however, shades of meaning and emphasis can be transmitted by varying the word order. These shades of meaning and emphasis are lexical functions in SB. Compare the two versions of the sentence "The two brothers loved each other very much": SB /ubhɔy bhratar pɔrɔm(ɔ) sɔmpriti chilɔ/ : CB /khub bhab chilo du bhaier/. According to Sen,[20] the initial and final expressions in the CB are the emphatic ones. In the case above, the SB compensates for this by the use of the strong terms /pɔrɔm(ɔ)/ and /sɔmpriti/.

3. The Bengali vernacular language, both from the point of view of prestige and the point of view of literature, was in a low state at the beginning of the nineteenth century. Bengali pandits much preferred to write in Sanskrit; when, as at Fort William College, they had to write Bengali, they rationalized it by saying that at least Bengali was "the best of the vernaculars, because of the preponderance of Sanskrit in it."[21] It can be gathered from this, I think, that the Bengali which these pandits wrote was not the Bengali which was the spoken language of the common people.

Strangely enough, the only people who were paying much attention to the vernacular at that time were the missionaries. It was the custom not many years ago to give the missionaries perhaps more credit than they deserve for the revival of vernacular Bengali and for the part which they played in raising the vernacular to a position of dignity. In the same way, it is the fashion now to debunk the idea that the missionary role

in the revival was significant. I think that it cannot really be questioned that their role was a significant one, but I think it is also true that this significance was not necessarily the result of a profound concern with the beauties of the language. There were certainly among the missionaries men who had a genuine love for the language and its literature.[22] Even had this not been so, however, the nature of missionary activity would have necessitated becoming involved in language and literature. These men were in daily contact with the people of Bengal; they had to learn the language as it was spoken, in order to carry out their work. Even if only from the point of view of text-books and propaganda, and again, in many cases it was far more than this, they had to learn to speak and write the language popularly, clearly, and usefully. The chief end of the missionaries in their use of the language was, as De says, "not to display their erudition", but to make themselves intelligible. These direct and utilitarian ends, together with the "tendency of the missionaries toward realism, directed the whole movement (of the Bengali revival) toward simplicity and naturalness."[23]

From this point of view, one of the major landmarks in this revival of the vernacular was the opening of Fort William College in Calcutta in Calcutta on May 4, 1800. The purpose of the College, according to Lord Wellesley's Minute, was "the better instruction of the Junior Civil Servants of the Company."[24] Actually, the College succeeded in a purpose far more significant in the long run than this. It contributed substantially to Bengali literature through the writings of its faculty and students, and through its emphasis on the vernacular as a language as worthy of study as the classical languages. The writings of the College faculty were published mainly at the Baptist Mission Press at Śrīrāmpur, which printed not only the original works of many Bengali scholars and writers, but also for the first time many of the classics of Bengali literature as well, including the Rāmāyana of Kṛttivāsa and the Annadāmaṅgal of Bhāratcandra.[25]

On the faculty of Fort William College, perhaps the most important single individual was William Carey, an English missionary who had gained an appointment as instructor in Bengali. Carey, together with other instructors at the institution, wrote large numbers of books in that language. These

books are, as De puts it, "far from being invulnerable from point of view of literary merit". Yet, despite mistakes in idiom and expression — perhaps inevitable for people writing a foreign language, no matter how familiar — the books which these men produced were for the most part simple, direct, and written in the language of the people.

Perhaps it would be wise to pass over Carey's Dialogues, the literary value of which seems to be questionable,[26] and turn for illustration to his Itihāsa-mālā, the literary value of which, I think, is not. The Itihāsa-mālā is a collection of some 150 stories taken from the Pañcatantra, the Hitopadeśa, and other Sanskrit sources, translated into simple Bengali prose. A few lines will indicate the style of the work.[27]

> ekajana ghaṭaka brāhmaṇa arthāt bibāher yojaka eka baner madhya diyā āsitechila | se sthāne eka byāghra ei ghaṭaka brāhmaṇke mārite udyata haile | brāhmaṇa bhīta haiyā krandana karite lāgila |

Here, except for the oral folk literature, are the first traces of literature in a style other than the ornate one which had become the norm among the pandits. The language of the Itihāsa-mālā can hardly be called modern colloquial Bengali. But I think the breadth of the gulf between it and the SB of the pandits will become clear when the above passage is compared with this one sentence from Mṛtanjaya's Prabodhcandrikā:[28]

> akārādi kṣakārāntakṣaramālā yadyapi pañcāśat saṃkhyakā kimba ekapañcāśat kimba saptapañcāśat saṃkhyā parimitā hauk tathāpi etābanmātra katipaya varṇāvalīvinyāsviśeṣavasatah vaidik laukik saṃskṛt prākṛt paiśācādi aṣṭādaśa bhāsā o nānādeśiya manuṣya jatīya bhāsāviśeṣavasatah aneka prakāra bhāṣāvaicitra śāstrato lokataḥ prasiddha āche

The difference is such that I think that even one entirely unfamiliar with Bengali will be able to see it. It is true that the subject matter of the two passages is quite different. But the difference is deeper than that. As De says,[29]

> "the genius of Sanskrit is not the same as the genius of Bengali; it would be a mistake to suppose that Sanskrit syntax should rule syntax in Bengali. ... The use of long drawn-out compound words, the occurrence of unusual phrases, and extensive borrowings from Sanskrit make it difficult for the

> unitiated to comprehend the sentences [of this SB] at a glance.

The last observation, to my mind, is rather an understatement of the case; but, in general, De's statement is an apt description of Mṛtanjaya's writing, which in turn is a good example of the SB of the pandits. It was not that Mṛtanjaya and the other pandits could not write, and write well, in a simpler fashion. Quite to the contrary.[30] They simply preferred the more ornate and complex SB constructions.

Why the pandits should have felt this way is a matter of speculation. Some reasons seem obvious. These were learned men, and Sanskrit the learned language; therefore, the closer to Sanskrit they could get, the more comfortable they felt. As is well known, the study of Sanskrit is for the pandits not a means to an end, but the end itself.[31] It is also true that these groups of learned men had but little contact with the masses, and had, therefore, little knowledge of and contempt for the means of the masses of expressing themselves. Finally, there might have been among the pandits, as there is today among certain people, a trace of the view of language held by the Pūrva-mīmāṃsā school of philosophy. According to this system of thought, the relationship between the word and its meaning is natural, and not created by convention. Language itself is an aspect of the Eternal Truth, and must be treated accordingly.[32] A corollary of this is the view that for serious subjects, a language of "spiritual adequacy" is necessary. This is an idea, not dissimilar to some modern American literary criticism, which is still current in Bengal. This view of the necessity of a "spiritually adequate" language, together with the idea that the subject matter of literature must be of a certain standard of moral quality, that literary themes must be edifying to be of value, results in the use of Sanskrit or of highly Sanskritized Bengali for any serious or creative form of literature. There is another view (not my own) of why the pandits reacted so strongly against the CB. The CB of the time included a great many Persian and Arabic words. The pandits, as Hindus, went to the other extreme and incorporated as much Sanskrit as possible into their Bengali texts. This is a view held by some scholars in East Pakistan. I am further told that the fact that there were many Perso-Arabic words in the

CB of the time is evidenced by the work of two of the earliest writers of the CB, Ṭekcānd and Vidyāsagar. The latter, usually thought of as a Sanskrit scholar, wrote under a pseudonym pamphlets in the CB directed against the critics of his child marriage and widow-remarriage ideas.

As can be seen from the case of the missionaries and of Fort William College, from at least one point of view the conflict between the two forms of Bengali began as a result of the cultural contact of Europe and Bengal. There was no conflict until the vernacular had been raised in status by missionary activity, by original writing, and by making available in printed form many of the standard and classic works of the Bengali language.

The contact of cultures produced another situation which had far-reaching effect on the language and literature of Bengal. Except for the missionaries and military, the European element in Bengal was primarily an urban-centered commercial element, manifested in the form of the Company. With the growth of the city of Calcutta as the focal-point of this commercial activity and its by-product of employment for Bengalis, the complex of Bengali society was substantially changed. And with the societal structure changed, of course, its language and literature. Not only were there new needs for the language — administration, commerce, and the like,[33] which necessarily affected its nature, but the new urban environment seems to have been particularly nourishing to vernacular literature. Calcutta grew rapidly not only as a commercial center but as an educational center as well. Cultural activity began to come to the fore. As a result of these things and others, the creation of a written literature became for the first time a professional activity, the function of an urban cultural stratum. A literary class, a class of professional writers, was produced. This was the first result.

The second was that these writers were interested in a kind of literature new to Bengal. They soon found that the traditional literary Bengali, the SB of the pandits, was not adequate for their new literary ideas. These new literary ideas were primarily prose style and the novel form.

It is well known that the first literature of any language is poetic literature, if for no other reason than that poetry is

easier to memorize and recite than prose would be. Although there were prose writings in Bengali before the coming of the Europeans, [34] prose was a form perhaps more fully developed in English. Prose was also a form which appealed greatly to the Bengali creative imagination. Thus, English prose became the model which the Bengali writers followed, in both style and subject matter. One of the first novels written in Bengali was a social novel. [35] The very concept of a social novel represented a break with traditional Indian literary ideals. This, together with the relative realism implied by such a literary form, led to a reconsideration of the value of the SB as a means of expression.

Furthermore, the educational and other institutions which were established by the British in Calcutta were naturally designed on familiar British lines. Instruction in such institutions was naturally from a British point of view. This gave rise to a large group of middle and upper class educated Bengali people who were oriented primarily toward the culture and literature of Europe, and particularly of England. The literary values of the consumers as well as those of the creators had been changed. How profound this change of values was can be easily seen by the most superficial examination of the Bengali literature of the mid-nineteenth century. The most striking case in point, I suppose, is that of Michael Madhusudhana Datta, a colorful and in many ways immensely tragic man living between two worlds. [36]

As Chowdhury points out in his paper in this volume, these changes had their effect in language as well as literature. Both SB and CB as they are known today are products of urbanization. Although a form of SB existed before the growth of Calcutta as an extremely Sanskritized Bengali, there was no such thing as CB before the city. The Bengali language then consisted of the SB and the regional dialects. The CB is the language of the educated middle and upper classes of Calcutta. Probably, as Chowdhury indicates, it is based upon the growth of the city. For the reasons noted above, SB tended to become less Sanskritized as a result of the cultural change. It moved away from the exclusive possession of the pandits and came into the possession of the new writer class, with their new literary aims. Looking at it in this way, which is perhaps

attempting to analyze it more than the situation warrants, since clear lines cannot always be drawn among the various divisions, it might be said that there have been five distinct forms of the language:

(1) The regional dialects, which of course, remain essentially unchanged throughout these various developments.[37]

(2) The CB as it was a century ago, a strictly spoken language.

(3) The CB as it is today, remaining primarily a spoken language, but being used increasingly for literature. As Tagore has pointed out, the CB is gradually taking on more and more SB elements.

(4) The old form of the SB, the paṇḍiti-bhāṣā, of which the Prabodh-candrikā of Mṛtanjaya (cited above) is an outstanding example.

(5) The modern-day SB which, influenced by cultural change and by the increasing use of the CB in literature, is far less definable grammatically than it was a century ago.

That there is a striking difference between the paṇḍiti-bhāṣā and the SB of a not much later date can be easily shown. Compare for example the passage from the Prabodh-candrikā above, written in 1813, to these lines from Bankim-candra Caṭṭopādhyāy's Rajsiṁha (1878):[38]

> eka prācīnā katakaguli citra becite āsiyā tāhāder hāte paṛiyāchila | hastidantanirmita phalke likhita kṣudra kṣudra apūrba citraguli | prācīnā bikrayābhilāṣe eka ekakhāni citra bastrābaraṇa madhya haite bāhira karitechila |

Bankim-candra's novel was written after the new concepts of prose and the novel had taken hold. But compare this in turn with a paragraph from a modern writer:[39]

> amalāke āj dekhte āsbe | pātrer nām aruṇ | nām śunei amalār bukṭite yeṇa arun ābhā chaṛie gela | kalpanāy se kata chabi na ākale | sundara suśrī yubā — baliṣṭha māthāy ṭeri gāye pāñjābī — sundar supuruṣ |

Even on the most superficial level it can be seen that the movement is toward simplicity and directness. The quotation from Bonaphul is almost pure CB.

Buddhadeva Bose says that after Bankimcandra prose "continued to be written in a manner artificially removed from the spoken language".[40] Although this is a true statement when the writings of Bankim-candra and his followers are viewed in the light of modern Bengali prose with its abundance of CB, it is also true that Bankim-candra represents the beginnings of the movement away from the unreadable stiffness of the paṇḍiti-bhāṣā. He was quite conscious of the artificiality of that form of the language.[41]

Although the main stream of Bengali literature in the nineteenth century continued to be that of the SB (although no longer paṇḍiti-bhāṣā), there were islands in the stream. A contemporary of Baṅkim-candra, Kālīprasānna Sinha, wrote in the CB, and wrote, according to Bose, "brilliantly". But whether it was due to his premature death or whether it was due to the fact that the literary world was not ready at that time for too radical a departure from tradition, the writing of Kālīprasānna Sinha remained merely an island in the SB stream.[42] It was, in fact, not until Pramatha Chowdhury, in the early twentieth century, that the assault of the CB on the citadels of literature began in earnest. In his journal Sabuj-patra Pramatha Chowdhury wrote scathing and lyrically bitter critical articles against the accepted norm of the SB, upholding the CB as a form of the language worthy and suited for adoption by the literary people of Bengal. A few rough translations from his articles will illustrate well enough the trend and expression of his thinking:[43]

> "We are the plaintiffs in this [literary] court ... we are only trying to reclaim our lost inheritance. Our opponents [i. e., the defender of the SB] know that possession [of literary control] is nine points of the law. ... They believe that the demand of our mother-tongue is but temporary, and that it will fade away, in time ..."
>
> "Once a custom is established, it remains through sheer inertia. This is the only explanation there can be for why things are as they are. In most people's minds, the fact that something is current is proof enough that it is necessary that it remain current. ...

> There are those who count their language as an inferior thing, and feel that writing literature in it is merely "the noisemaking of vulgar people ..."
>
> "If there is no introduction of the human element into literature, it is not literature. ... Artificiality has no place in art ..."
>
> "In Bengal, little boys (and others) think that if you put an anusvāra on a Bengali word, you have Sanskrit; or that if you cut the anusvāra or visarga off a Sanskrit word, you have Bengali. If you cut the tail off a monkey, will he become a man? ..."

The voice of Pramatha Chowdhury was a strong and an effective one. But it was with Rabindranath that the CB found its place in literature. It is well beyond the scope of this paper to go into the development of Rabindranath's style. That has been covered well elsewhere.[44] Let it be said only that from his earliest days, Rabindranath had a deep feeling for the colloquial language. The letters which he wrote from Europe at the age of 18, published under the title Yurop prabāśir patra, are in pure CB[45] — as Bose says, "lively and lovely prose, the prose of everyday speech". In terms of his more formal work, it was not until his novel Ghore bāire that his CB style came to full realization. The novel was written in 1901-2. I think it is fair to say that from that time, the CB began to usurp the place which had once been occupied by the SB alone.

"The end of the story has not yet come; but there will be a time when the younger will sit upon the throne alone."

Notes

1. Leonard Bloomfield, Language 48-53 (New York, 1933).

2. See for example Preserved Smith, The age of the reformation 451 ff. (New York, 1920); E. de Moreau et al., Histoire de l'eglise Vol. XVI, 94ff.

3. S. J. Al-Toma, The teaching of Classical Arabic to speakers of the colloquial in Iraq; a study in linguistic duality. Doctoral thesis for the Graduate School of Education, Harvard University, 1956; unpublished.

4. The father of a friend of mine recalls instances of pandits and scholars of his generation speaking in SB in ordinary circumstances. My friend himself has never heard this.

I am told that in an article by Bankim-candra on the contribution of Ṭekcānd's prose to Bengali literature, instances of such SB speech are recorded and are held up to ridicule as pretentious and laughably absurd. I have not had access to the article in question.

5. It is not unlikely that, if one were to follow out these parallels, it would be found that a closer comparison to the Classical - colloquial Arabic distinction in some respects would be that between Sanskrit and Bengali.

6. The regional dialects represent, of course, the third major division of the language. The regional dialects are very highly divergent in some cases — as between, for example, Birḅhum in western West Bengal and Chittagong in what is now East Pakistan. Although the question of these regional dialects does not relate directly to the present problem, it is interesting to note that the dialects too have had their influence on the form of the standard spoken language, and thus, indirectly, on the literary one. For example, the form /sathe/, "with, accompanying", although originally an East Bengal dialect form, is now commonly used in the CB throughout West Bengal. It is now at least as common as the once more usual form /sɔŋge/.

7. See Buddhadeva Bose, An acre of green grass 20 (Calcutta, 1948). This was true in Bengali despite the fact that even in classical Sanskrit drama the characters spoke in Prakrit when appropriate. See A. B. Keith, A history of Sanskrit literature 30-1 (Oxford University, 1920).

8. Among the writers of CB who preceded Rabindranath are Kālīprasānna Sinha and Ṭekcānd. See Sukumar Sen, Bāṅgālā sāhitye gadya 94 ff., 86 ff., (Calcutta, 1935). According to M. Chowdhury, however, the work of these men might be better described as "including colloquialisms" than as being "written in the colloquial language". For a brief discussion of Rabindranath's contributions, see p. 58 below.

9. Bāṅgālā-bhāṣā-paricaya 39 ff. (Viśvabhārati, 1938).

10. Ibid.

10a. In this paper phonemic transcription is enclosed in slant lines; other Bengali material in roman letters is in transliteration.

11. The reason for this change, Tagore says, is not only that class distinctions are becoming less and less clear-cut, but also that with the increase in education and general awareness, the subject matter of every-day discussion is changing. This means that terminology which was once limited to the SB, since it was considered that only SB terminology could be used for serious literature and discussion, is now entering the CB.

This concept, that serious or abstract thought cannot be expressed in the CB is, as we shall see, not uncommon in Bengal. Buddhadeva Bose would approach the matter from the other direction, and say that common and homely ideas and images cannot be expressed in the SB.

12. The somewhat sarcastic name for this mixture of CB and SB is guru-caṇḍāli. The term indicates people high and low on the social scale respectively.

13. The list was devised in Poona in the summer of 1955 for the purpose of comparative study of the Indo-Aryan languages, and not for the purpose for which I have used it here.

14. See Sukumār Sen on Baṅkim-candra in Bāṅgālā sāhitye gadya, 107 ff.

15. The data shows that the form /bristi hɔoa/ is also common in the CB.

16. /ɔpekkha kɔra/ is also a common form in CB. These are examples of the optional category mentioned above. As before stated, the movement toward the center is from both directions. A great many SB forms may be used in the spoken CB, and, depending on the circumstances, almost any CB form can be used in writing the language.

17. Cf. Bloomfield, Language 492 ff; S. K. Chatterji, Bāṅgālā bhāṣātattver bhumikā 50 ff., (Calcutta, 1950).

18. For a treatment of compounds in Bengali, see Sukumār Sen, Bāṅgālā bhāṣār byākaran 140 ff. (Calcutta, 19).

19. Ibid.

20. Ibid.

21. From Mṛtanjaya's Prabodh-candrikā, quoted in S. K. De, History of Bengali literature in the 19th century 278 (Calcutta, 1919). The passage reads:

> anyānya deśīya bhāsā haite gauṛadeśīya bhāsā uttamā sarbottamā saṃskṛtabhāsā bāhulya hetuka

22. De (op. cit. 127) says of William Carey that he "found the appointment (as instructor in Bengali at the College) bringing in its train responsible duties ... and affording him an opportunity to cultivate 'the beautiful language of Bengal'". The source of the quotation is not given.

23. De, op. cit. 289.

24. Minutes in Council at the Fort William, by His Excellency the Most Honourable Maquis of Wellesley, dated August 18, 1800. Quoted in De, op. cit. 120.

25. The press at Śrīrāmpur printed works not only in Bengali, but in many other languages. In the appendix to the 10th Memoir (1832), it is shown that two hundred and twelve thousand volumes in forty different languages had been issued between 1801 and 1832. See De, op. cit. 157, note 1, and J. C. Ghosh, Bengali literature 100 ff. (Oxford University, 1948).

26. The Dialogues are called in Bengali Kathopakathā. The literary merit of the work is in question seemingly because it was not written with a literary end in view. As a matter of fact, Carey himself did not write most of it, but simply "employed some sensible natives to compose dialogues upon subjects of a domestic nature, and to give them in precisely the natural style of the persons supposed to be speakers". (From Carey's introduction to the work; see D. C. Sen, History of Bengali language and literature 725 (2nd ed. Calcutta, 1954). Thus it is considered by some merely as an interesting sociological commentary.

27. The full text of the story is given in De, History 148. Translation might read:

> Once there was a Brahmin match-maker (ghaṭaka), a marriage-agent, who was coming through a forest. In that forest there lived a tiger, who was ready to eat this ghaṭaka Brahmin. The Brahmin was very frightened, and began to weep.

28. Quoted by De, History 217.

> Even if there were fifty or fifty-one or fifty-seven letters of the alphabet, there could only be so many different arrangements of the letters. There are eighteen forms of language — Vedic, ordinary, Sanskrit, Prakrit, Paiśāci (i. e., demonic speech), etc. Among men, the characteristics of speech vary according to their several countries and castes. All the various kinds of language are well known from the śāstras.

29. History 220.

30. Compare, for example, this passage from the same work, the Prabodh-candrikā quoted in De, History 221):

> bāṭīr nikaṭe giyā āpana strīke ḍakila... tāhār strī gatikriya kahila ogo āmi yāite pāribo nā āmār hāt yorā āche |
>
> Going up to his house, he called his wife ... His wife slowly answered — I cannot go. My hands are joined in supplication.

Even superficially it can be seen how much simpler the

language is. The most striking characteristic of this passage is its lack of compounds.

31. See M. B. Emeneau, India and linguistics, JAOS 75. 145-6 (1955).

32. See S. Radhakrishnan, Indian Philosophy Vol. II, 389 ff. (2nd ed. London, 1951).

33. A parallel to this in some ways is the origin of Hindustani, which was also developed for military and commercial purposes. In the case of Hindustani, it should be noted, the impact was of a non-European Muslim culture.

34. See Sen, Bāṅgālā sāhitye gadya, 1-10, and De, op. cit., Appendix I, 455-86. According to many scholars, the earliest specimen of Bengali prose is the Śūnya-purāṇa of Rāmāi Paṇḍit (published by the Baṅgiya-sāhitya-pariṣad, Calcutta, 1912). S. K. De, Haraprasād Śāstri, and others, find in this work "traces of the original writings of Rāmāi Pandit, going back to at least the 13th century A. D.". The question of the date of the work, and in fact the question of whether or not it is really prose, are still open. There is hardly space to deal with them here.

35. Ālāler gharer dulāl of Ṭekcānd Thākur (whose real name was Pyāricānd Mitra), written in 1858.

36. Discussions of this man and his life and works can be found in any literary history of Bengal in English, and in any of several biographies in Bengali. In English, see for example J. C. Ghosh, Bengali literature 137-48.

37. See S. K. Chatterji, Origin and Development of the Bengali Language, Calcutta University 1926, Vol. I, 138-146.

38. In the Basumati edition of Baṅkim-candra's works (Baṅkimcandrer granthābalī), Vol. I, 3. Translation:

> An old woman had come to sell some pictures, and had fallen into their hands. The pictures were exquisite miniatures painted on ivory tablets. The old woman, eager for a sale, brought out a picture from underneath her garments.

39. From Bonaphul's Śreṣṭha galpa (Calcutta, 1957), the story called Amalā. Translation:

> Today he will come to see Amalā. The bridegroom's name is Aruṇ. At the mere sound of his name, it was if the glow of dawn (aruṇ) was spreading in Amalā's breast. How many beautiful pictures she painted in her imagination of him — handsome, gracious, young,

strong, his hair dressed, a pāñjābī on his body — a beautiful man.

40. Bose, An acre of green grass 19.

41. See Sen, Bāṅgālā sāhityer kathā, 173-4.

42. Bose (op. cit. 19) describes it in this way:

> Kālīprasānna Sinha ... in his Nutam pañcar nakṣa, first brought literary grace to the spoken tongue. If he had lived reasonably long, the dispute between the two styles might have been settled long ago. But he died unpardonably young, and Baṅkim-candra was the only author worthy of imitation.

43. From the essay entitled Sādhubhāṣā bānama calit bhāṣā in the collection Prabandha saṃgraha Vol. I, 318-27 (Viśvabhārati, 1952). The feeling about the closeness of the relationship between Sanskrit and Bengali mentioned in the last passage is still widely held in Bengal. Actually, lexical considerations aside, Bengali is probably one of the Indo-Aryan languages farthest removed from Sanskrit. See Chatterji, Origin and Development of the Bengali language.

44. Sen, Bāṅgālā sāhitye gadya, 157-205.

45. Published under the title Yurop prabāśir patra; see Sen, Bāṅgālā sāhitye gadya 192 ff.

THE LANGUAGE PROBLEM IN EAST PAKISTAN

Munier Chowdhury

Dacca University

0. Introduction

1. Linguistic background

2. Dialects of East Pakistan

3. Phonological variation

4. Morphological variation

5. Present trends

0. British India was partitioned in 1947 and two new independent states came into being. One of them is the Islamic Republic of Pakistan, whose two zones lie geographically more than a thousand miles apart. The eastern half of it, which has a little over half the total population of Pakistan, is known as East Pakistan. It also happens to be the eastern half of the former Province of Bengal whose other half now belongs to the Republic of Bharat. We thus today have Bengali spoken in two immediately contiguous lands belonging to two independent states with many conflicting interests. There is, of course, a significant distinction between the Bengalis of the two republics: East Bengal Pakistanis are overwhelmingly Muslims, whereas West Bengal Indians are predominantly Hindus. On the other hand, the language of East Pakistan is not understood by the people of West Pakistan. The latter, although belonging to various linguistic groups, have shown a general inclination to accept Urdu as the language of culture education and administration.

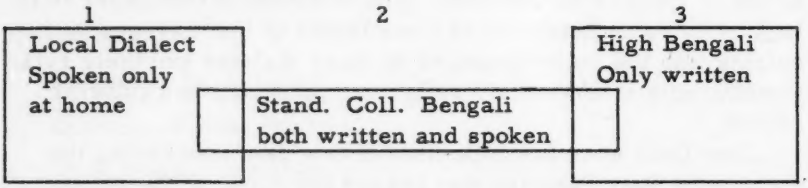
This paper is primarily about Bengali as it is spoken today in East Pakistan. The total area of East Pakistan is about one-sixth that of the whole of Pakistan, but it has a large variety of dialects, some of them even mutually unintelligible. In the

very near future a new standard colloquial language with a special prestige value may pattern out a blend of these dialects. If and when the bond with West Pakistan with its different set of Indo-Aryan languages becomes more intimate, coupled with a greater zeal for Islam, the Bengali of East Pakistan may develop its potentialities in a direction quite different from that of West Bengal. There is also the spirit of nationalism in the country which devoutly wishes to replace English by Bengali in as many spheres as possible. The intention of this paper is to explain in some detail the various facets of these issues by inquiring into the major features of these dialects and their relationship with other social, religious, political, and cultural values.

Any field work for collection of new data concerning the dialects of East Pakistan was beyond the scope of the present study. The broad facts of the local dialects were culled chiefly from the Linguistic Survey of India,[1] which offers copious texts from different dialects of Bengal. General background information and material on the current situation come from the author's personal experience. His special advantages for this may be summarized: (1) native speaker of a highly divergent local dialect, Noakhali, (2) long residence in Dacca, the capital of East Pakistan, (3) extensive travel throughout Bengali speaking territory as a result of his father's being a civil servant, and (4) active participation in the life of the University, where the language problem is perhaps felt most acutely. Another important aid in the writing of this description of the Bengali of East Pakistan was the unpublished doctoral thesis of C. A. Ferguson on the grammar of Bengali.[2]

1. An educated Bengali in East Bengal (Pakistan) as well as in West Bengal (India) is fully aware of three distinctly different kinds or styles of Bengali, the use of which in various socio-cultural spheres is rigidly restricted by well-defined traditions. One does not use one's local dialect (unless it happens to be one of or around Calcutta) in the office, University or in any discussion tinged with formality. It is used only within the home or informally outside. The second is the Standard Colloquial Bengali (SCB) which has developed out of the local dialects spoken in and around Calcutta, the former capital and industrial and cultural center of unpartitioned Bengal. This is

a standardized form which, like the dialect of Tuscany for Italian,[3] was accepted by all Bengalis as the elegant mode of speech of all Bengalis in all formal circumstances outside the home. Another significant feature of this colloquial form is that it can also be used in writing. The third kind is Standard Literary Bengali or High Bengali (HB), which is only a literary form and is never used colloquially. One might describe the state diagrammatically as:



The stratification is also graded in terms of prestige value. The local dialect, when widely removed from the patterns of the SCB such as any East Bengal dialect may be, is looked d down upon as rude and vulgar. It has been used by popular writers of fiction as a means to evoke laughter. The rural population, who form ninety percent of the total and are mostly agriculturists, know almost nothing of the correct forms of SCB. They use the local dialect in all circumstances. SCB is the form one must adhere to in polite company; it is the form the radio and the cinema communicate through; it has prestige even as a written form. High Bengali (HB) rates even higher. If the topic is grave, warranting a scholarly discussion, even the SCB is rejected in favor of the HB. Only in recent years has there been a shift in emphasis, with SCB usurping a major share of the glory that formerly was exclusively bestowed on HB.

The relationship between HB and SCB is the subject of a paper by E. A. Dimock in this volume;[4] emphasis will be given here to the characteristics of the local dialects of East Pakistan and the consideration of certain general problems and present trends.

2. Accepting the possibilities of other kinds of variation and other kinds of interrelationships among the variants, we can for the purposes of our discussion tentatively follow Grierson's classification[5] of the dialects of East Bengal (coinciding approximately with present-day East Pakistan) into four main groups.

1. Northern Bengali.	The dialects of Dinajpur, Rajshahi, Bogra, Pabna.	
2. Rajbangshi.	The dialects of Rangpur.	
3. Eastern Bengali.[6]	The dialects of	(a) Dacca, Mymensing, Tippera, Backergunge, Sylhet, Haijong.
		(b) Faridpur, Jessore, Khulna.
4. Southeastern Bengali.	Chittagong, Noakhali, Chakma.	

In the absence of any systematic study regarding mutual intelligibility among these dialect groups it is difficult to state precisely how they stand in relation to each other.[7] We can only generally describe it as follows: (1) dialects of the fourth group are progressively less understood as we move upwards from it to the first group (where they become almost unintelligible); (2) the dialects of the first three groups may be understood by all people all over East Pakistan including speakers of the fourth group; (3) Sylhet of the third group, insofar as it shares closely the characteristics of the fourth group, has also approximately the same area of mutual intelligibility. Haijong of the third group and Chakma of the fourth group are spoken by two extremely small groups of people and are not easily understood by even the speakers of the nearest dialects belonging to the same group. The latter has a writing system all its own, widely different from Bengali orthography.

Let us now examine some of the significant features that differentiate these dialects from one another.

In phonology, as shown below in 3., there is very little deviation from the inventory of phonemes of SCB, a fair amount of deviation in the phonetic actualization of certain phonemes in Eastern and Southeastern Bengali as opposed to all other dialects, and considerable difference in the distribution and frequency of certain phonemes. On the whole, however, apart from a few striking features of pronunciation there is relatively little variation in phonology.

In morphology the picture is different. It is true that the basic structure of the word is about the same in all dialects, but the actual form of many derivational and inflectional suffixes varies enormously from one dialect to another, and this is probably to be regarded as the most important area of divergence either between these dialects and SCB or among the dialects themselves. The facts are summarized in 4.

The dialects of East Bengal do not vary very much in syntactical patterns except in certain negative and interrogative construction. In Chittagong-Noakhali dialects the negative particle na may be used before the verbal form: /nɔ zaiyum/ 'I, we will not go' as opposed to SCB use after the verb /ǰabona/. Example of the interrogative ki 'what': /nam ki tɔr/ 'What is your name?" as opposed to SCB /tomar nam ki/.

It has not been possible to make a systematic investigation of the lexical differences among the dialects, but it is the author's feeling that this will not prove as important as morphological variation.

The phonemes of SCB may be listed as follows:[8]

k	kh	g	gh	ŋ		i	u	ĩ	ũ
č	čh	ǰ	ǰh			e	o	ẽ	õ
ṭ	ṭh	ḍ	ḍh			æ	ɔ	æ̃	ɔ̃
t	th	d	dh	n			a		ã
p	ph	b	bh	m					
r	ṛ	l	š	h					

Some problems in the analysis of Bengali phonology are: (a) whether in addition to an alveolar or palatal /š/ we should also posit a phonemic dental /s/; (b) whether ḍ and ṛ are two phonemes or variants of one; (c) whether bilabial aspirated consonants are phonetically stops or fricatives; (d) the status of the semi-vowel y; and (e) the status of aspiration. They are mentioned here because interestingly enough they also represent significant points where dialectal variation is conspicuous.

One more thing. Both Muslims and Hindus use SCB and both have absorbed certain influences of Persian (through this of Arabic) which reigned as the language of the court of Bengal and India for six hundred years. There are today about 2500

words of Perso-Arabic origin in Bengali, and a few suffixal forms like -giri, -dar, used by monolingual speakers of Bengali. In general the Persian sounds have been adapted to the phonological structure of Bengali. But such adaptation has not been so thorough with Muslims as with Hindus. The tradition of teaching children how to recite the Holy Qoran correctly is still very widely practiced. Also, Muslim names in Bengal are chiefly Arabic names, and many words relating to the practice of Islam are Arabic. One definite feature of "Muslim phonology" is the fact that many Muslim speakers of SCB seem to make consistent distinctions between s and š and between z and j. It is difficult to find absolute minimal pairs, but these four sound types in the speech of such Muslims cannot be accepted as contextual variants of two phonemes. In this kind of SCB [s] and [z] occur intervocalically and initially before a vowel. Frequently occurring examples include asia, samad, zaman, azad, aziz. All these words the vast majority of Hindus would pronounce with [š] and [ǰ]; we may therefore include these two as phonemes /s z/ of SCB as spoken by the average educated Muslim Bengali.

Specific phonological differences among the dialects of East Pakistan may now be taken up.

(a) NORTHERN BENGALI

Phonemically it has the same system as SCB. Some oddities in a series of lexical items indicated that in some instances NB has /r/ for SCB /r/, /l/ for SCB /ṛ/, and in Malda /s/ for /š/. In vowels from the point of view of SCB there is a confusion between /o-ɔ/ and /æ-e/.

Examples:	NB	SCB	
	/sorile/	/sorire/	(in the body)
	/pæṭ/	/peṭ/	(belly)
	/bɔrɔ/	/boṛo/	(big)

(b) RAJBANGSI

Phonemically the same. Examples of deviations from SCB: Confusions between /s-čh/, /l-n/, /z-ǰ/, /d-dh/, etc. and /o-ɔ/. Also occasional elisions of certain consonants.

/ase/	/ačhe/	(is)
/zela/	/jela/	(District)

/randhite/	/rā̆dte/	(to cook)
/gɔrbo/	/gɔrbhɔ/	(pregnancy)

(c) EASTERN BENGALI

The distinction here is in the very phonemic system. Eastern Bengali has no /č/, /čh/ or /s/, and no /ǰ/. Instead it has / ¢ /, / ¢h / (affricates of the ts variety) and /z/. It has also a glottal stop. The series of voiced aspirated stops does not occur.

Sylhet has fricativized variants for both /k/ and /p/ with a greater frequency of occurrence than the non-fricative variant.[9] They seem also to be in free variation with /kh/, /ph/. Thus we have [k], [p], medially, when the fricative counterpart occurs in the preceding syllable or when it is geminate; [x], [ɸ] elsewhere. In vowels it has no /æ/. Like Chittagong-Noakhali it possibly has consonantal /y/ and /w/.

With reference to SCB pronunciation of words, EB shows further confusions between /r-ṛ/, /ɔ-o/, /æ-e/. Initial /s/, and sometimes medial /k/ of certain SCB words, is dropped or replaced by /h/. On the other hand, initial SCB /h/ in EB is not often pronounced, and medially is replaced by a glottal stop. These characteristics in ordinary conversation immediately betray the identity of the speaker as one belonging to EB.

Dialects of Faridpur-Jessore-Khulna constitute the link between East Pakistan Dialects and SCB. Geographically and linguistically they are nearer to Central West Bengal in the order mentioned. Where SCB has /ch/ Rajbangsi has [s], and for /ǰ/ it has [z], but /č/ is not [¢], nor are /s/ and /h/ lacking as in other EB dialects.

(d) SOUTHEASTERN BENGALI

The phonemic characteristics of EB are here more conspicuously observable. When compared with SCB the deviations are found to be thorough and complete. There is a marked divergence in vocabulary also. At least that is what one suspects from the radically different phonetic appearance of words. /k/ seems to be spirantized. Grierson notes its similarity to the Arabic voiceless velar spirant. The consonantal /y/ and /w/ are distinctly tense and cannot be compared with the non-syllabic front and back vowels, which are relatively lax.

4. In this section some examples of the dialectal variants of the suffixal morphemes will be given. They are stated in terms of the SCB norm.

(a) NORTHERN BENGALI

Nominative -e has in the plural /-te/: /čhaote/ 'children'. Objective plural suffix -der, -digɔke has /-dek/: /čakɔrdek/ 'servants'. A very unusual form of dative termination is /-gune/: /amakgune/ 'to me', /bapakgune/ 'to father'.

The Locative -e has the alternants /-et, -ɔt/: /kaštet/ 'in wood', /dešɔt/ 'in country'.

In verbal forms, the participial -e (HB -iya) has the variants /-æ/, /i/ after a stem-final consonant or vowel respectively. Ex: /kɔræ/ 'having done', /pai/ 'having received'.

The personal ending in the 1st. per. future is /-im, -am, (-mu)/ instead of /-bo/: /bolim/ 'shall say', /ǰam/ 'shall go' etc.; cf. SCB /bolbo/, /jabo/. The 3rd per. has no vowel finally: /hoil/ 'it became', /čhil/ 'it was'; cf. SCB /holo/, /chilo/.

In bordering areas near Malda /-gor/ for genitive plural and /-mu/ for 1st per. fut. may also be found.

There are isolated instances of adverbial or loc. -e being /-i/: /samni/ 'in front', /asti/ 'slowly'.

(b) RAJBANGSI

Nom. pl.: /-gula, -gila/, gen. pl. -(e)r plus /-ghɔr/: /balɔkgula/ 'boys', /čakɔrerghɔr/ 'of the servants'.

Obj.: /-k/ as in NB.

Genitive /-kar/: /balɔkkar/ 'of boy'

Loc. /-ɔt, -ot/: /balɔkɔt/.

Verbal forms:

1st pers.	present and perfect	/-ð/: /ačhð/ 'I am'
	past	/-nu/: /ačhinu/ 'I was'
	future	/-m, -mu, -mo/
2nd pers.	past	/-lu, -llu/: /kaillu/ 'I did', 'you did'
	future	/-bu/.

(c) EASTERN BENGALI

First the forms that are generally found in this group are mentioned, then a few specific references are made.

For SCB obj. suffix /-ke/, EB has /-ere, -re/: /bapere/ (to

father). For SCB gen. /er/ EB has /ɔr/, gen.-obj. pl. is /-go/: /ɖakɔrgo/ 'to servants', /tago/ 'their'. In Mymensing obj. plural has /-rare/: /dusrare/ 'to friends; in the locative /-ɔ, -t/: /dilɔ/ 'at heart', /gɔlat/ 'on neck'.

Verbal forms:

1st per. fut. /-mu/: /komu/ 'I shall say'.

In Mymensing/-am/. In the 3rd per. past the endings are (Mymensing) /-lɔ, -l/.

For SCB participal -e we have /-a/: /koria/ or /koira/ 'having done'. In Tippera it is added after boudling of the final consonant of the stem: /boilla/ 'saying', /uiṭṭa/ 'getting up'.

The infinitive in Tippera is /-to, -tam/ instead of SCB /-te/: /bɔrto/ 'to fill'. /koitam/ 'to say'. In Mymensing it is /-ɔt/: /bɔrɔt/ 'to fill'. Infinitives for forming inceptive compounds show even greater contrast when /-bar/ is used for SCB /-te/: /aigaibar laiglo/ 'began to approach'. In Sandip /-ɔn/ is used: /karɔn dilo/ 'began to do'.

In Haijong:

Nom. /-ra, -a/ : /pɔlara kɔe/ 'the child says'

Obj. /-ge, -ra/ : /polarɔge thɔle/ 'she placed the child' /takaradi/ 'give the rupee'.

Gen. /-lak, -la/ : /rajalak/ 'of the king' /ranila/ 'of the queen'

Loc. /-mi, -ni, -mini/ : /dešɔni/ 'in the country'.

As regards verbal forms, Haijong, like all the other dialects, uses the SCB forms freely but also has "a set of forms for the various tenses which do not change for the number or person" possibly "due to the influence of Tibeto-Burman Languages".[10]

Example:

Past	/-bar, -ban/
Fut.	/-iŋgɔ/
Infini.	/-bake, -bak/
Imper.	/-marek, -mar/

In Sylhet,

Gen. sing.	/-ɔr/ : /gɔrɔr/ 'of homes'
Loc.	/-t, -o/ : /gɔrɔt, gɔro/, 'in house'
Nom. plu.	/-ain/ : /gɔrain/ 'houses'
Past 2nd	/-lai/ : /dekhlai/ 'you saw'

3rd /-la/
Fut. 1st /-mu, -bam/
2nd /-bai/
3rd /-ba/
Conditional /-tam, -tai, -ta/ (past)
" /-ain/ (present, 3rd, honorific)
Progressive present
1st /-tram, -yar, -ram/ : /zaitram/ etc. 'I am going'.
2nd /-trai, -tre/
Imperative /-un, -uka/

The infinitive of purpose or desire has the 'most noticeable construction' in Sylhet. The ending (SCB /-te/) changes for person /-tai, -to, -ta/ :

/tumi zaitai ɕao/ 'you wish to go'
/he zaito ɕae/ 'he wishes to go'
/tain zaita ɕain/ 'he (hon.) wishes to go'

(d) SOUTHEASTERN BENGALI

Nom. /-e/ is very common.
Abl. /-tun/
Gen. /-ɔr/
Loc. /-ɔt/
Plural /-gun, -un/ : /kʊ̃urgun/ 'dogs'
Present tense
1st /-r/ : /khair/ 'I eat'
2nd /-ɔs, -ɔr, -or/
3rd /-er, -tan/
Past tense
1st /-(g)yam, -gi/ : /korgyɔm/ or /kariyam/ or /korgi/ 'I did'
2nd /-(g)yɔ, -la/
3rd /(-g)yɔ, -l/
Future tense
1st /-(g)yum, -bɔ, -bam/
2nd /-ba/
3rd /-be/
Imperative /-ɔ/, when negative /-io/ : /nɔ korio/ 'do not'
Honorific /kɔrɔtak/.
Participial /-yare/ : /koriyare/ 'having done'.
(in Chakma /-inai/ : /ǰeinai/ 'having gone')

To take a final example from Chakma:

	Present	Past	
	Sing. & Plur.	Sing.	Plur.
1st	/agi/ 'I am'	/eluŋ/ 'I was'	/elaŋ/ 'we were'
2nd	/agɔs, ne/	/ele/	/ela/
3rd	/age, ne/	/el/	/elak/

To summarize: the morphological peculiarities of the individual dialects of East Pakistan are most striking in three features: a) the pronominal stems and their related forms; b) case endings of nouns and pronouns; and c) the tense-morphemes and personal endings of verbs.

Example (a)

"I" is /ami/ in SCB, /mui/ in NB, /ăi/ in Noakhali (SE), /mɔe/ in Chakma (EB). "Of me" is /amar/ in SCB, /mor/ in NB, /ma-lagik/ in Haizong (EB), /mɔ/ in Chakma (SE), /ăr/ in Noakhali. "Of us" is /amader/ in SCB; /hamar/ NB; /amrar/ Mymensing (EB); /morgo/ Barisal (EB); /mogo/ Dacca (EB); /amalak-beden/ Haijong (EB); /ărar/ SE; /amarar/ Chadme (SE).

Example (b)

	SCB :	/čakɔrder/	'to servants'
	NB :	/čakɔrdek/	
	EB :	/ɕahɔrgo/	
	SE :	/ɕhaorgun/	
or,	SCB :	/gɔlae/	'on neck'
	EB :	/gɔlat/	

Example (c)

SCB :	/bolbo/	'(I) shall say'
NB :	/bolim/	
EB :	/bolmu/	
SE :	/koiyum/	

In offering a general description of the morphological peculiarities of these dialects of East Pakistan we should take into account the fact that most of these dialects may also use forms that are virtually identical with those of SCB, freely in alternation with forms that do not occur in SCB.

5. It is clear from the preceding section that for many East Bengalis there is a co siderable difference between local

dialect and SCB. As a result, the people of East Bengal, in particular the speakers of Eastern and Southeastern Bengali, always suffered one disadvantage: they tried hard but never could satisfactorily wield this tool of SCB in normal verbal traffic. It was not the grammar that bothered them most, for the rules were simple and few and one more or less got to know them along with a general education. But when they grappled with it in actual utterance they floundered miserably, and the more wary they were the more pitiful became their lapses. Girls, who may be more eager in attaining grace and elegance both in personal appearance and in social behavior, fared better in the struggle. The men lagged behind. Even when their grammar was correct in terms of SCB their pronunciation of individual items was often deplorable, their intonations too dialectal. Socially this had been very compromising.[11]

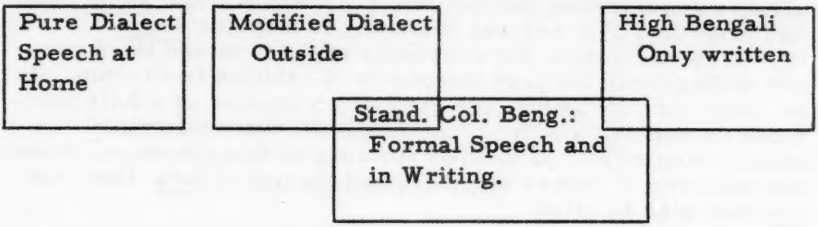
But the situation today is different. Now SCB can also be looked upon as a standard form of Bengali spoken in Indian West Bengal. The spirit of antagonism against an invidious distinction among dialects is thus reinforced. Literally there are fewer people around in East Pakistan to jibe at you for your frailties. A sense of release from constraining obligations in verbal behavior is noticeable everywhere. Intellectuals are found engaged in discussing economics, philosophy, art, politics all in a language not definitely the same old SCB. When boy meets girl today at the university they freely begin to exchange innuendoes in dialectal forms which a few years ago would have acted as a veritable antidote to any embryonic romance. Even the broadcasting house is now tolerant of many 'talkers' who satisfy their listeners enormously in spite of many deviations from SCB in pronunciation.

What strikes one most in the situation is the absence of any one standardized spoken form which everyone aims to cultivate. It is not true that everyone is using his own dialect with a vengeance in and outside the home. The necessities of modern living are bringing about an incessant commingling of diverse dialects of many tiny groups of people living close together and are also creating the hankering after a common convenient standardized form. There is a conscious effort to choose forms that are common with the listeners' dialect or that are presumably widely understood, an effort to reduce the idiosyncrasies of

one's own dialects, and an attempt to incorporate the easily imitable allied forms of the SCB as well as HB within the frame of this naturally emerging standard colloquial Bengali of East Pakistan. The past tense /hoečilo/ 'it was' is not being replaced by such forms as /(h)oisilo/ (Dacca), /oisil/ (Noakhali), /oisele/ (Bar.) or by HB /hoiačhilo/ but by something that strikes a mean between SCB and HB: /hoiʧhilo/. Both SCB and HB influence it in retaining the initial /h/ but the modification of the affricate is characteristic of EP. Similarly, the first person future ending is commonly /-mu, um/ instead of SCB /-bo/.

For the gen. -r you may have /-go/; /-re/ instead of SCB /-ke/, /-er/. /-e/ may be used oftener in the nominative. Pronominal /tara/ may take over /ora/ 'they'. Phonemically the chances of remaining very close to SCB are very high except in the area of the affricates. The deviations from SCB will be still conspicuous particularly in words containing /r, ṛ/ or /ɔ, o/. All these portend to one thing: the Dacca-Mymensing group of dialects may play the major role in moulding the Standard Colloquial of East Pakistan. As the capital city grows, becoming the center of economic, social, political and cultural activity of the whole of East Pakistan, its dialect, too, will gain in gravitational force. The SCB based on the dialect of Calcutta and its surroundings may be conjectured to have taken shape the same way.

The whole situation in EP is still in such a transitional, formative stage that all predictions remain open to question. Moreover, its very transitional nature has added more complexity to the interrelationship among the already existing styles of Bengali. The new form in East Pakistan is not well defined or codified and as such hardly acts as a substitute for the old SCB which is still the form to be adopted in all strictly formal circumstances such as the radio, the classroom, the pulpit. No EP dialect or any blend of EP dialects as yet can be used in writing. Hence, making the burden worse, the educated EP speaker has now four instead of three models of Bengali to juggle with. They may be diagrammatically shown as:


Pure Dialect Speech at Home
Modified Dialect Outside
High Bengali Only written
Stand. Col. Beng.: Formal Speech and in Writing.

/ai zaiyum/ /ami ǰamu/ /ami ǰabo/ /ami ǰaibo/ 'I shall go'

/mui zaibam/

/ami zamu/

Notes

1. G. A. Grierson, Linguistic survey of India. Vol. V. Calcutta, 1903.

2. C. A. Ferguson, The phonology and morphology of Standard Colloquial Bengali. Univ. of Pennsylvania Ph. D. thesis, 1945.

3. Rabindranath Tagore, Bangla Bhasha Porichoy, Calcutta, 1938, p. 47.

4. E. A. Dimock, Literary and colloquial Bengali in modern Bengali prose.

5. Op. cit., p. 19.

6. Note that the territory covered by the linguistic term "Eastern Bengali" does not coincide with the political entity East Pakistan.

7. The Bengali dialect situation is similar, on a smaller scale, to that of Hindi as characterized by John Gumperz, Language Problems in the Rural Development of North India, The Journal of Asian Studies, 1957, Vol. XVI, No. 2.

8. Cf. S. K. Chatterji, A Brief Sketch of Bengali Phonetics BSOS, London, 1921, Vol. II, Part I; C. A. Ferguson and M. Chowdhury, The phonemes of Bengali, Language 35:4 (1959).

9. This comment is based on personal observation; it is not clear from Grierson's material.

10. Grierson, op. cit.

11. On the other hand the dialects of East Bengal had some effect on the forms of SCB. Tagore pointed this out in 1938

giving as illustration the fact that the postpositional šathe 'with' had crept into SCB and was beginning to displace šɔŋge.

In verbal forms, the alternants of the first and third person endings with the past morpheme -l- shifted from -lum, -le to -lam, -lo. Possibly, the gradual acceptance of a fully aspirated pronunciation -čh- of the progressive morpheme -č- as equally elegant SCB is another instance of this influence. Certain features of phrase structure and the use of tara 'they' for ora may also be cited.

SOCIAL DIALECTS IN DHARWAR KANNADA

William McCormack

University of California, Berkeley

0. Introductory — background and procedure

1. Differences, Brahmin : non-Brahmin

2. Differences, backward classes : other classes

0.1 In Dharwar, and in the adjoining districts of Belgaum and Bijapur, there appear to be three dialects of conversational Kannada which correspond to the three main cleavages in the social system. The three social classes to which there is a corresponding class dialect are the Brahmin, the non-Brahmin, and the Harijan. These three social levels are known popularly as the 'advanced', 'middle', and 'backward' classes, respectively. The principal caste represented by the advanced classes in the three districts are the Deshastha Brahmins, who belong either to the Smarta (Saivite) sect or to the sect of the Madhvas, who are primarily worshippers of Vishnu. Deshastha Brahmins number about 2.5% of the population of Dharwar, Belgaum and Bijapur districts. The total population of these three districts, which were called collectively the 'Bombay Karnatak' before their merger with the new Mysore linguistic province on 1 November, 1956, was reported in 1951 to be 4,698,479. The middle classes of the three districts are mainly followers of the Lingayat, or Virasaiva, religious sect, whose members form about 20% of the population in the new Kannada-speaking linguistic state. Lingayats are concentrated in their greatest strength in Dharwar, Belgaum, and Bijapur districts, where they account for about 35.5% of the population. The two principal castes of the Lingayat sect are everywhere the Jangams, or priests, and the Banajigas, or traders. The backward classes, whose hereditary occupations are held to be ritually unclean by traditional Hinduism, include about 18% of the population in the

new Mysore state, and these classes are not concentrated in any one or few districts. According to the 1931 Indian census, backward classes (Harijans) form 8.0% of the population of Dharwar, Belgaum, and Bijapur. But a significant proportion of the people practicing 'unclean' occupations in these three districts are Telugu speakers. In 1951 1.3% of the combined population of these three districts were returned as Telugu-speaking, and their hereditary occupation is, by inference, that of sweeper. The principal Kannada-speaking castes which fall into the 'backward' classification are the Madigas and the Holeyas.

0.2 Social dialects are recognizable in colloquial Kannada only, and not in the literary dialect of so-called 'educated speech'. Just as the literary dialect is essentially uniform for the three major regional dialects of Kannada, namely, the Dharwar, Bangalore and Mangalore types, so also the spoken graphic forms are the same for speakers of different social levels in the Dharwar region. A Kannada speaker requires about seven years of formal schooling before he can understand the literary dialect completely, and some practice or additional years of formal education may allow him to master it as a spoken style. Statistics given by the 1951 census show that 1.25% of the population in Dharwar, Belgaum and Bijapur districts had undergone seven years or more of formal education. But even among educated people, the spoken graphic forms constitute an 'artificial' style, which is mainly limited to the stage, to formal oratory, and to pedantically-inclined individuals. Section 0.5 below describes some of the phonemic features of 'educated speech'.

0.3 A preliminary study was made to find out with what degree of accuracy a small sample of the population of Dharwar district could recognize the social level of a speaker from about five minutes of his tape-recorded monologue. A large panel of twenty-eight primary school teachers listened to seven speakers of different social levels, and a small panel of nine college students listened to twelve speakers. Five entries were to be filled in for each speaker: his caste, native place, age, education, and family income. The tapes were edited of content clues to the identity of the speaker, and they were made by the standard procedure of asking the speaker to imagine family situations and to speak to relatives or friends on homely subjects. The

caste of the speaker was correctly identified by the panels 40% of the times, misidentified 35% of the time, and there was 25% no response to the caste question. The three Brahmins of the Dharwar region were among the speakers most often correctly identified. One of these Brahmins scored 87% correct identification, and as his speech was full of literary borrowings, we might infer that his high score was due to his confirming the social stereotype that 'Brahmins speak like books'. The four backward class speakers were almost never correctly identified, and the reason for this was that they spoke a mixture of Brahmin, Lingayat and literary forms which effectively masked their own distinctive class dialect. Some urbanized backward class speakers regularly use such a mixture when they speak before members of other social classes, but it is a fact which most Dharwar people do not recognize because they automatically assume that a backward class speaker will use his own distinctive linguistic forms. The best recognition score for a backward class speaker was 25%, two scores were 0% and one was 10%. There were no 0% scores for other than backward class speakers. The best recognition score, 89%, was made by the smaller panel on a Lingayat speaker, but there was a content clue in the form of a Lingayat personal name which was not edited from this tape and this probably accounts for the high score. The small panel scored 78% identification of another Lingayat and 56% on a third. The large panel scored 48% identification of one Lingayat speaker and 32% on the other Lingayat whom they listened to.

0.4 This paper will attempt to present some of the morphological differences in the speech styles of Dharwar Brahmins, Lingayats, and members of the backward classes. The comparison is based on about an hour's tape-recorded material in which Brahmins were speaking in imaginary family situations and on forty minutes' tape-recorded conversation in a backward class family's house when only family members were present. (The recording machine was operated by a family member.) A Lingayat linguistic informant then listened to these two tapes, and the forms which he would say differently were noted. About 220 forms were marked as Brahmin-Lingayat differences, and fifty-two forms were marked as backward class — Lingayat differences. Lexical differences accounted for about 20% of the Brahmin-Lingayat differences, whereas only three lexical

differences between backward classes and Lingayats were noted from the taped conversation.

0.5 The possibility of differences in the phonemic structure of the three social styles of speaking Kannada in Dharwar is complicated by the problem of individual differences in the use of loanwords in different degrees of assimilation to the native phonemic pattern. The amount of material in the tapes on which the morphological comparisons are made in this study is not sufficiently large to handle this lengthier phonemic problem. Loanwords have come into Dharwar Kannada from English, Sanskrit, Hindi, Marathi, and probably from other languages as well. In general, it appears that loanword phonology in Dharwar is related to individual factors like education, intelligence and motivation to learn languages, and the opportunities given by the professional contacts of the speaker to learn languages. For example, villagers who raise cotton and so market it at the local cotton-trading centers have extensive contact with Bombay-type Hindi, which other villagers and townspeople lack. Anyone with close relationship to government affairs, including village officers, knows a certain amount of English, Marathi, and Hindi. A person living in a town has more contact with languages, as a rule, than the villager, and since the languages are taught in schools, the educated person generally knows more languages. The high rate of educated Brahmins relative to other classes, their traditional association with the Sanskrit language, their close association with government, their concentration in towns and cities, and the fact that intermarriage between Marathi and Kannada speaking families is more frequent among Brahmins all combine to increase the probability that a loanword will not be assimilated to the native phonemic pattern of a Brahmin speaker. But these appear to represent a combination of circumstances affecting individuals, as it were indirect consequences of Brahmin caste membership, in contrast to the distinctive grammatical style which all Brahmins use as a direct consequence of their caste. The phoneme /f/, in contrast to the /p/ substitution of uneducated persons, appears in loanword used by Brahmin speakers in the taped conversations which were used for the recognition of social level of speaker study (Section 0.3). But this seems to be more reasonably interpreted as a consequence of the English schooling of the

Brahmin speakers and of their contact with English in government service, rather than as a direct consequence of social level or caste membership. The /f/, /p/ contrast and similar phonemic information might be characterized as part of an 'English-educated speech' style. It seems doubtful, however, that there is a 'Sanskrit-educated speech' style, which would exist on the phonemic level apart from the (Kannada) 'educated speech' style mentioned in Section 0.2. This latter style incorporated phonemic distinctions between aspirate and non-aspirate stops, nazalized and non-nazalized vowels, and distinguishes three voiceless sibilants (a flat blade-alveolar /ś/, narrow groove /s/, and retroflex /ṣ/ which has an aspirate release as the second member of a consonant cluster when followed by a vowel), all of which are of recognized Sanskrit origin. These distinctions are not made in Kannada by uneducated persons, nor by those educated persons who regard them as 'artificial'. The Sanskrit palatal stop + homorganic nasal is rendered as /dny/ in the loanword 'knowledge' by educated persons, in contrast to the colloquial form of the word, /gya:na/.

1.0 Differences in the phonemic shape of suffixes which occur in identical morphemic environments with non-contrastive meaning are among the most important differences in the speech of Brahmins and non-Brahmins. One of these distinctive Brahmin suffixes is socially recognized as marking the isogloss which divides Brahmin from non-Brahmin speech. This is the verb suffix 3rd person-singular-neuter (i.) /ədəl/. There are also differences in the phonemic shape of verbal bases as Brahmins and non-Brahmins use them, and there are differences in the use of compound base in one style as against a simple base in the other. Kinship nouns of address and greeting suffixes have also been included in the discussion of the distinctive features of Brahmin colloquial style.

1.1.1 Differences in the phonemic shape of verbal suffixes indicating person-number-gender of the actor are illustrated by four examples in this section. There are two contrasting sets of person-number suffixes in which gender is distinguished in the third person (singular: male, female, neuter; and plural: human, non-human). There is a third contrasting set, the imperative, but this set lacks first-person-singular suffix, and

gender and number are not distinguished in the third person. In the examples, these sets will be marked as (i.), (ii.), and (iii.). Two allomorphs of the non-Brahmin suffix 3rd person-singular-neuter (i.) are shown, each of which was given as the non-Brahmin equivalent for the corresponding Brahmin form /ədə/. The Brahmin form is listed first in all examples.

/e/	; /æ/	1st person-singular (ii.)
/e:vi/	; /i:vi/	1st person-plural (i.)
/o:ṇa/	; /u:ṇu/	1st person-plural (iii.)
/ədə/	; /ayti/, /ə/	3rd person-singular-neuter (i.)

1.1.2 Four additional person-number (gender) suffixes for which there are Brahmin non-Brahmin differences as in Section 1.1.1 are listed below. These were discovered when the Brahmin forms listed in the preceding section were checked with Brahmin informants. The Brahmin suffix is listed first, followed by the corresponding Lingayat form. The more frequent allomorph in use by Lingayats of the morpheme 3rd person-singular-male (ii.) is /a/. The non-Brahmin allomorph /æ/ listed here as different from the corresponding Brahmin /a/ is found with about 36% of the verbal bases which non-Brahmins (Lingayats) use.

/e:ni/	; /i:ni/	1st person-singular (i.)[1]
/i:di/	; /i/	2nd person-singular (i.)
/a/	; /æ/	3rd person-singular-male (ii.)
/a:və/	; /a:vu/	3rd person-plural-nonhuman (i.)

There are fourteen (64%) person-number (gender) suffixes for which Brahmin and non-Brahmin forms are the same. Four are shared in set (i.), seven in set (ii.) and three are shared in set (iii.).

1.1.3 Relational suffixes may be added to verbal bases in order to form conditional particles in Kannada, which have the syntactic function of phrase modifiers. Three suffixes of this type, in which the Brahmin form differed in phonemic shape from the corresponding non-Brahmin form, were found in the tape. The pairs of suffixes are as follows, the Brahmin form listed first:

/o:tle/ ; /ku:dle/ 'soon after'
/likke/ ; /a:ka/ 'to' or 'for'
/go:ta/ ; /konta/ Durative action

1.1.4 There are locative suffixes in Kannada which are added to nominal bases to indicate the location of a noun with respect to the action of the verb. Two such suffixes may be added to the same nominal base, so that there can be base + suffix$_1$ + suffix$_2$, as in /mani.kaḍ.inda/ 'from the house'. The starred forms in the following list indicate Brahmin forms which non-Brahmins also use in some contexts, but in at least one context (the context in which the Brahmin used the suffix in the recorded conversation) the non-Brahmin would employ the second form which is listed here.

*/inda/ ; /æ:ga/ 'from'
*/kaḍe/ ; /la:ga/ 'at'
/tuttu/ ; /suttu/ 'around'
/oḷaga/ ; /a:ga/ 'in that'
/æ:ka/ ; /ige/ 'to'
/ukka/ ; /ige/ 'to'

1.2.1 One common difference between Brahmin and non-Brahmin colloquial styles is the use by Brahmins of compound verbal bases in environments where the other style has a simple base. The use of compound bases is felt to be closer to the literary norms of written Kannada. Examples of this style difference are given below. The Brahmin form is listed first, followed by the non-Brahmin equivalent, and then the form with a verbal suffix with which the Brahmin form occurs on the tape, and finally the non-Brahmin equivalent of this. Period indicates morpheme boundary in the examples.

/ma:ḍ.idd./ ; /ma:ḍ./ /ma:ḍ.idd.i./ ; /ma:ḍ.i/ 'you (sg) have done'
/bay.tir./ ; /bay./ /bay.tir.t.ara/ ; /bay.t.ara/ 'they will abuse'
/a:g.bit./ ; /a:g/ /a:g.bit.ədə/ ; /a:g.e:ti/ 'it is done'
/idd.ir/ ; /ir./ /idd.ir.beku/ ; /ir.beku/ 'it might be'

1.2.2 The phonemic shapes of seven verbal bases which occur in the Brahmin tapes are different when non-Brahmins use these bases in identical environments. Two pairs, /ku:/; /kun/ and /horḍ/ ; /hon/, 'sit' and 'start out', respectively, have different base forms for Brahmins as compared to non-

Brahmins in all environments. The remaining five examples have shared base forms which in some environments both groups use. In the following list, the shared base form is listed last.

/.go:./	; /.go:n/	bound reflexive verb	/.gonḍ./
/haḍ./	; /haḍa./	'gave birth'	/haḍa./
/ho:./	; /hu:./	'go'	/ho:g./
/a:y./	; /a:./	'become'	/a:g./
/idd./	; /ad./	'be'	/idd./

1.3 There are some generally acknowledged differences between Brahmins and non-Brahmins with respect to kinship nouns and other forms of address and greeting. Half a dozen of the examples given in this section occur in the tape, and, as non-Brahmins are quite familiar with these differences, the remaining examples were elicited from a Lingayat informant, and then checked with a Brahmin. The Brahmin forms and the non-Brahmin equivalents are listed below. Forms are starred to indicate that non-Brahmins also use these Brahmin forms, but in different contexts, which are explained in parentheses after the gloss. Differences in verbal suffixes for general greeting are listed first, and then kinship nouns are given. The Brahmin form is listed first.

*/.o/ ; /.apa/	greeting any man, especially with the verb base 'come', as in /ba:r.o/ ; /ba:r.apa/ (Non-Brahmins use /.o/ in greeting intimate friends and inferiors.)
*/.u/ ; /.be/	greeting any woman, with the verb 'come' /ba.O/ ; /ba:r.be/ (Non-Brahmins use /.u/ in greeting wife or any younger relative, friends, and inferiors.)
*/.u/ ; /.apa/	greeting any man, with 'come' /ba.O/ ; /ba:r.apa/ (Non-Brahmins use /.u/ in greeting younger relative or inferiors.
*/.ə/ ; /.evva/	addressing mother, with 'come' /ba:r.ə/ ; /ba.evva/, /ba:r.be/ (Non-Brahmins use /.ə/ to greet wife only.)
*/.ə/ ; /.ekka/	addressing elder sister (as above)
*/akka/ ; /ekka/	'elder sister' (Non-Brahmins use /akka/ as a kinship term of reference only, whereas Brahmins use it both as a term

of reference and as a noun of address.)

*/akka/ ; /evva/ 'mother' (As above for non-Brahmins; /akka/ is a noun of address 'mother' for Brahmins and /avva/ is the term of reference among both groups.)

*/aṇṇa/ ; /eṇṇa/ 'elder brother' (Non-Brahmins use /aṇṇa/ as term of reference.)

*/aṇṇa/ ; /eppa/ 'father' (As above for non-Brahmins; /appa/ is the term of reference among both Brahmins and non-Brahmins.)

/mama:vara/ ; /ma:va/ 'father-in-law', 'mother's brother', 'father's sister's husband', 'elder sister's husband', 'brother's wife's father'

/atti.avara/ ; /atti/ 'mother-in-law', 'father's sister', 'mother's brother's wife', 'brother's wife's mother'

*/adji/ ; /emma/ 'grandmother' (Non-Brahmins use /adji/ for 'great-grandmother'.)

*/muttæ/ ; /edja/ 'grandfather' (Non-Brahmins use /muttæ/ for 'great-grandfather'.)

1.4 There is a residue, constituting about 10% of the marked Brahmin-non-Brahmin differences, which appear to be syntactical differences and are still to be analyzed. It is not known which of these differences are due to borrowing from literary styles and which are genuine differences in the order and arrangement of morphemes. One syntactic difference is, however, apparent from the Brahmin tapes, and that is the use of the relational suffix /anta/ in a much wider range of contexts by Brahmins, as compared to non-Brahmins.

1.5 There is a larger residue, about 40% of the original list of forms which were marked as distinctive of Brahmins' style, that was rejected as showing 'no difference' after rechecking with the informants. Most of these spurious differences arose out of literary borrowings by Brahmin speakers. As the Lingayat listener was instructed to mark the differences between the taped conversation and his own colloquial speech, these forms were naturally included in the first list of differences. But when they were checked with other Brahmin

informants, these literary loans were discovered to have arisen from the momentary pedantic inclination of a speaker in the tape and were not in colloquial use among Brahmins. A few of the spurious differences were due to the mental set of the Lingayat listener, who had been instructed to find differences, and upon rechecking with him these differences were found to be illusory.

2.0 The majority of differences between the speech style of backward classes in Dharwar and the speech styles of the other two social classes appear to be differences in the phonemic shape of verbal suffixes which are used in the same morphemic environments as their equivalents in the other two speech styles. There are also differences in the use of compound verbal bases and in the phonemic shape of allomorphs of the same verbal base morpheme, as they are used by members of backward classes. There are some distinctive forms of address and kinship nouns, which only are to be found in the colloquial utterances of Harijans. But the lexical inventory, the vocabulary, of the Harijans' speech appears from this study to be practically identical with that of other non-Brahmins in Dharwar. The tendency of urbanized members of the backward classes to substitute forms which are in use among members of other social classes for their own distinctive forms was noted above in Section 0.3. This fashion has also included some forms which Harijans share with other non-Brahmins; especially the distinctive Brahmin 3rd person-singular-neuter /.ədə/ is favored by these urbanized speakers, rather than the non-Brahmin /.ayti/, which most Harijans use. Such borrowings as are made by these urban speakers will be classified along with the literary loanwords, and they will not be considered as part of the colloquial style of backward classes.

2.1.1 The distinctive person-number (gender) suffixes in the tapes of backward class speakers and their equivalents in non-Brahmin (Lingayat) style are given in this section. The suffix used by backward class speakers is listed first. The set to which each suffix may be assigned is noted as in Section 1.1.1 by the figures in parentheses, (i.), (ii.) and (iii.).

/e:nə/	;	/i:ni/	1st person-singular (i.)
/a:ḷu/	;	/a:ḷa/	3rd person-singular-female (i.)
/uḷ/	;	/ḷu/	3rd person-singular-female (ii.)

/a/ ; /æ/ 3rd person-singular-male (ii.) (64% Lingayat bases take /a/ rather than /æ/; cf. 1.1.2.)

/e:vu/ ; /i:vi/ 1st person-plural (i.)
/u/ ; /vi/ 1st person-plural (ii.)

2.1.2 One additional person-number (gender) suffix which is distinctive of backward class speech in Dharwar was discovered when the forms given in Section 2.1.1 were checked with Harijan informants. It is listed below, with accompanying Lingayat form.

/ur/ ; /ru/ 3rd person-plural-human (ii.)

Backward class speakers share, then, fifteen (68%) person-number suffixes with the middle class Lingayats, and they share ten (45.5%) of these suffixes with Brahmin speakers. The Brahmins and backward class speakers share one common difference from Lingayat speakers, as their 3rd person-singular-male (ii.) form is /a/. But this is not a major shared difference, since among Lingayats /a/, as the more prominent allomorph of 3rd person-singular-male (ii.) morpheme, occurs with three times as many verbal bases as the variant form /æ/.

2.1.3 There appear from the taped conversations to be three frequently used relational suffixes which have distinctive phonemic shapes for backward class speakers. The Harijan form is listed below first, followed by the Lingayat form, which is also the form used by Brahmins in Dharwar.[2]

/kæ:sa/[2] ; /O/ (/O/ is phonemically zero)
/nandre/ ; /re/
/ut/ ; /ta/

2.1.4 There are two examples in the recorded conversations of locative noun suffixes, which have a distinctive phonemic shape for backward class speech. One is /i:gi/, which is used as /ige/ by Brahmins and Lingayats. The second example is /a:g/, which only Harijans attach to personal nouns of address (names) and to kinship nouns. The Brahmin and Lingayat equivalents of /a:g/ distinguish two situations, where the backward class speaker would apply /a:g/ to both. A Lingayat uses

/an/ (Brahmin form is /anna/) when the action of the verb is directed toward the noun to which the suffix is attached, and he uses /ug/ (/əgə/ is the Brahmin's form) if it is intended to associate the noun with the action of the verb.

2.2.1 There are a few examples of Lingayat-backward class differences with respect to compound verbal bases. Three examples from the tape are given below. The backward class form is listed first.

/ma:ḍ.id./	; /ma:ḍ./	/ma:ḍ.id.e:nə/	; /ma:ḍ.i:ni/	'I did'
/.al.ikk./	; /.idd./	/he:ḷ.al.ikk.illa/	; /he:ḷ.idd.illa/	'didn't say'
/a:/	; /a:k.hat./	/hel.a:.tidda.ri/	; /hel.a:k.hat.tidda.ri/	'were telling, sir'

2.2.2 Three examples are listed in this section of Harijan and Lingayat verb bases which differ in phonemic shape in identical environments. A form of the base which both groups share is listed last.

/.go:/ ; /.gon/	bound reflexive verb	/.gond./
/gont/ ; /go:ṭ/	'know'	/go:ṭ/
/a:/ ; /a:k/	'become'	/a:k/

2.3 Examples of distinctive Harijan greeting forms and kinship nouns are given below, followed by their Lingayat equivalents. The verbal suffixes of general greeting are listed first, and then the differences in kinship nouns.

/.a:ko/[3] ; /.apa/	greeting any man
/.o/ ; /.apa/	greeting any man (The suffix /.o/ is also used by Brahmins, but with voice-qualifier 'authority', whereas Harijan usage may frequently involve the voice-qualifier 'fear, trembling'; Lingayats use this suffix in greeting intimate friends and inferiors.)
/aḷiya.deva.ru/ ; /.apa/	addressing son-in-law
/bi:g.ru/ ; /.apa/ = males /.be/ = females	addressing affines
/ekka/, /eppa/, /emma/ etc. ; /akka/, /appa/, /amma/ etc.	

reference terms. (The backward class speaker uses the non-Brahmin nouns of address in both referring to and addressing relatives. Lingayats have a separate set of nouns for address and for reference.)

2.4 No syntactical differences were marked by the Lingayat informant in about half a dozen repetitions of the tape of backward class conversation. There may be one or two differences in the use of stress and intonation morphemes, but further checking and analysis is needed on this point.

2.5 The residue of 'imaginary' differences which were marked from the backward class tape by the Lingayat listener was about 10% of the list of differences. These spurious differences were due to the mental set of the listener to find differences and not to the marking of literary borrowings, which were rare in this tape.

Notes

1. In Belgaum the Brahmin form is /i:ni/.

2. In the Bijapur style of speaking the /kæ:sa/ suffix is not limited to the speech of Harijans as it is in Dharwar and Belgaum.

3. In the Bijapur District this usage is not limited to backward classes.

FORMAL AND INFORMAL STANDARDS IN THE HINDI REGIONAL LANGUAGE AREA

John J. Gumperz

University of California, Berkeley

C. M. Naim

University of California, Berkeley

0. Introduction

1. The Social Function of the Standard Stratum and its Relation to Local Dialects

2. Linguistic Variation Within the Standard Stratum

3. Conclusion

0. The Hindi regional language area is the part of North India, where Hindi serves as the official medium of local government and administration. This area, which includes the present states of Rajasthan, Delhi, Himachal Pradesh, Uttar Pradesh, Madhya Pradesh, Bihar and parts of Punjab, is not only the largest of the linguistic units on the subcontinent, but also stands apart in other respects. In other sections of the country regional languages are limited to compact, geographically well defined areas, a fact which has made possible the recent organization of linguistic states. Furthermore there is in each area a well recognized standard, known by a single name, which although often linguistically distinct from local dialects, has served as the prestige form for some time. Literature in this standard is generally read and accepted by the educated, including those who know English well. All this is not the case with the Hindi area. Some forms of the modern standard language have been current since the early days of Muslim rule, but many distinct varieties exist. These varieties are spoken in geographically widely separated areas, used

in different social situations, under differing prestige conditions and are called by many different names.

We find a multitude of conversational styles — apart from the local dialects — variously called Hindustani, Khari Boli, Dakkhini Hindi, Dakkhini Urdu, Bazar Hindustani, or also Urdu and Hindi.[1] They are spoken not only within the regional language area itself but throughout urban North India in centers as far apart as Karachi, Bombay, Lahore, Calcutta, Ahmedabad, Hyderabad and many others. The exact spread of all these forms of speech has never really been surveyed and there is furthermore little agreement on the application of the above names. Different individuals may give different names to the same variety while others use identical terms to refer to linguistically distinct varieties. On the literary level two distinct styles are usually recognized: Urdu, which takes its learned vocabulary largely from Persian and is written in an adaptation of the Perso-Arabic script and Hindi, which draws on Sanskrit for its loans and employs the Devanagari script. There is however much stylistic diversity, even within these two named entities. One school of Hindi writers has introduced so many Sanskrit borrowings that its writings are incomprehensible to a speaker of the colloquial language trained only in Urdu. A second group writes in a way which is very close to the colloquial and which could become Urdu merely by changing script and substituting a few key terms. Other writers have introduced a number of local dialect features into their works thereby causing difficulties for readers from distant areas.[2] There is furthermore some controversy about the application of the term Hindi to the various medieval and post medieval literatures. The tendency in recent years has been to use Hindi for all texts produced in the present regional language area. But, many object to this practice and continue to employ the old dialect names. Literary Urdu is a little more uniform than Hindi but writers still vary greatly in the extent to which they borrow from Persian.

The bulk of the linguistic studies so far have concentrated on one or another of the above named entities and have as a rule treated it as a separate unit with little reference to the speech community in which it was used or to related forms of the language. We have grammars of Hindi based on writings

in the Devanagari script which often include a number of the sub-regional dialects.[3] Analyses of Urdu may deal either with the literary style or with one or another of the spoken varieties current among the educated groups.[4] Of the grammars which carry the title Hindustani, some cover the style current in the Indian Army — others the so called Bazaar Hindustani or a somewhat Urduized urban speech style.[5] Even Grierson's Linguistic Survey of India, which contains the most detailed information on speech distribution operates with unit dialects or styles.[6] It is moreover not based on direct field work, but rather on written information elicited by a somewhat circular procedure in which informants were asked to furnish translations of a standard text into one or another named speech form as known from local sources. These dialects were then again treated as separate units. Studies like the above, while useful for those whose interest is confined to literature or to the structure of a single variety, nevertheless represent rather arbitrary abstractions from actual linguistic behavior, which are of little aid in clarifying the bewildering range of varieties of Hindi-Urdu found in North India. If we are to understand speech distribution, we must describe it in terms of the society in which it operates.[7]

The present study therefore takes as its unit of analysis not a particular style, but all forms of the standard language in the regional language area as defined in the first paragraph. We will disregard the traditional language names and will feel free to deal with any standard varieties used in this area, excluding however others not found there such as the Urdu of Pakistan, or the bazaar Hindustani of Bombay and Calcutta. The paper is divided into two parts, part one describes the social environment of the standard, its relation to local dialects, its development and function within the speech community. Part two presents a linguistic analysis of speech variation and its relation to other factors in the social system.

1. The total range of speech varieties found among urban and rural, educated and uneducated, high class and low class residents of the Hindi regional language area may be described in terms of three strata, each corresponding to its own network of communication and having different functions in the social system. The dialects spoken at the local or village level form

part of a chain of mutually intelligible varieties of Indo-Aryan stock, extending through all of North India from the Arabian Sea in the West to Assam in the East. They vary geographically from village to village and socially from caste group to caste group. [8]

Superposed above this chain we find the sub-regional dialects, varieties which avoid the most divergent localisms and are understood over somewhat wider areas. They are the native speech of the traditional service and trading castes, residing in small bazaar towns as well as in the larger urban centers, who cater to the needs of the rural population. Villagers employ these dialects as a second speech style for inter-group or inter-village communication. Grierson's classification based on historical similarity sets up for the present Hindi area five groups of varieties called 'languages' each of which in turn contains several dialects. These groupings are: Western Hindi, Eastern Hindi, Rajasthani, Bihari and Pahari (further subdivided into Eastern, Central and Western Pahari). [9] Only the first two of these groups show a direct relationship to Hindi, others are linguistically closer to neighboring regional languages. Furthermore, as with the local dialects, language distance increases with geographical distance, so that there is mutual unintelligibility among speech forms at opposite extremes of the region.

Before the advent of the British, in Muslim times, many of the sub-regional dialects had great importance as the literary media for the then dominant devotional and court poetry. The social function of these literatures was quite limited, however, and different from that of a modern standard. We have no evidence for the existence of a single prestige style or range of styles which served for all types of writing and formal discourse as is the case in modern western societies, but rather there was a series of quite distinct styles, each used for a particular purpose. Administration was carried on in Persian or Sanskrit, Vedic was used for traditional Brahminical rituals, trading communities had their own special medium of communication and there seem to have been almost as many co-existing different literary styles as there were distinct religious cults or literary traditions. Each of these literary styles in a way served as the symbol or trade mark of a particular group. Compositions were intended primarily for oral presentation and were spread

beyond the original dialect area through the religious establishments, migrant preachers, ascetics and performers belonging to that group. Thus Braj Bhasha was the literary vehicle for the poets of the Krishna cult of which Surdas and Mirabai are the best known, Avadhi was the language of Tulsi Das's Ramayan, and Maithili is famous for the Vaishnavite poetry of Vidyapati. Literary activity of this type now lives on only in oral folk tradition and has all but lost its former social importance. The medieval literary dialects however still receive a great deal of scholarly attention. Moreover some of the better known, e. g. Maithili, Avadhi, or even Grierson's historical constructs, Rajasthani and Bihari, continue to be listed among the major languages of India.[10]

The third stratum is that of the standard, which we shall refer to by the collective term Hindi-Urdu. The varieties of this stratum are in turn superposed above the local and sub-regional dialects. They are the native speech of only a minority of urban residents; for most people they serve as a second or third speech style. Modern Hindi-Urdu had its origin in the speech of the Gangetic Doab, just north and northeast of Delhi, in the districts of Meerut and Moradabad. Soon after the Muslim conquest, a trade language, based on this speech, became accepted as the lingua franca of the courts, army camps and trading centers of the new ruling groups. The use of this language for trade purposes spread through much of North India, into many districts where the native dialects were quite different, as far as the coastal district of Bengal and Gujerat and Maharashtra. Some hundred years later we find records of several literary styles bearing various degrees of similarity to the spoken trade idiom. Among these are the Khari Boli of Amir Khusrau and later of Kabir, the Rekhta of Delhi, the Dakini of the Muslim courts of the Bahmani kingdoms in the Deccan, where the local population spoke a Dravidian tongue, and several others. Throughout this early period these styles were similar in function and distribution to the sub-regional dialects described above. Compositions were confined to a particular form of religious or court poetry, which was not the only, but one of many coexisting literary traditions in the area. As a matter of fact, during the seventeenth and eighteenth century Khari Boli poetry was largely replaced by Braj Bhasha.

The development of the modern standard language did not really begin until the second quarter of the nineteenth century, after the consolidation of British rule, when literary Urdu replaced Persian as the official language for the purpose of local administration. After that time a new Urdu prose style developed; the language came to be taught in schools and greatly gained in prestige. Because of its association with Islam, however, the new style was not universally accepted. While many Hindus were on the whole willing to use Urdu for the purpose of business and administration, few accepted it as a literary language. A second prose style was created by discarding the Persian alphabet in favor of the Devanagari script and substituting Sanskrit borrowings for the Persian loan words. The new Hindi for some time became the symbol of the rising forces of reform Hinduism as represented by the religious centers of Benaras and Allahabad, the Arya Samaj, and the Brahmo Samaj movements. Religious reform and Hindi education went hand and hand and a widespread system of Hindi vernacular schools and colleges was set up paralleling the Urdu school system. By the end of the nineteenth century the two new standards had all but replaced the sub-regional dialect literatures throughout the present Hindi area. But they did not spread into all regions where the Hindustani bazaar language was employed. In Gujerat, Bengal and Bombay other literary styles were already well established while spoken varieties of Hindi continued to function as trade idioms with little prestige.[11] After independence in 1947, Hindi replaced Urdu throughout the present regional language area as the sole official medium of administration and education and is now learned by all. Urdu literature however continues to flourish.

With the increasing tempo of urbanization and the recent improvements in communication, the varieties of the standard stratum are beginning to make greater inroads among the local dialects. Native speakers of Hindi-Urdu still represent a minority, although an ever increasing one. They are found primarily among the westernized, urban groups, who have moved away from their home region and have lost connection with their rural past. For a much larger number of individuals the standard continues to function as a second or third speech style, used only in certain social situations, e. g. on formal

occasions, with individuals of different social background, in the office, in school, on the college campus etc. But the number of situations which call for the use of the standard is growing and sub-regional dialect forms are being more and more confined to the family circle.

The existence of the standard as a common prestige form has a profound effect on the region as a whole. It makes it possible to overcome local variations in speech and culture. A person who knows Hindi or Urdu not only has little trouble in making himself understood wherever he travels in the area, regardless of local dialect, but he is also certain to find many other individuals who read the same literature and in general share his intellectual interests. Even among those who barely know it the standard has a certain prestige position. We find it used for example in public meetings with audiences made up entirely of uneducated monolingual villagers. In the traditional Muslim 'Milad' the Maulvi is required to preach in a highly persianized Urdu and in certain Hindu meetings of the 'Kirtan' type the audience expects a very sanskritized form of Hindi. The function of the standard on these occasions is primarily ritual since very little concrete information is conveyed.

The relation between the standard and the sub-regional and local dialects may be likened to that between Redfield's great and little tradition.[12] The former serves both as a model for prestige imitation and as a channel of communication through which new concepts are funneled into the other two levels. Linguistically these innovations are reflected in loan words and phonological and morphological borrowings, leading to the reinforcement of those local dialect features which are similar to the standard and to the creation of differential rates of linguistic change on two sides of a regional language boundary. Traditional dialect surveys which tend to concentrate on old or antiquated speech, disregarding newly borrowed features show little evidence of the above processes. But if we were to make a modern study of every day speech, say at three equidistant points, A, B, C, along our North Indian chain of mutually intelligible local dialects, where A is separated from B and C by a regional language boundary, we might expect the overall language distance between A and B — measured in terms of the totality of grammatical, lexical and semantic differences — to

be greater than that between B and C. Speakers in the latter two locations should share a number of learned and elegant features, borrowed from the regional language, which we would not find in A. The standard thus acts as a unifying force linguistically as well as culturally, a common mold which counteracts diversity at the local level and separates the regional language from other neighboring districts.

If we consider the group of standard speakers from the point of view of communication, we find that it has a number of characteristics which set it off from the rest of the population.[13] Those who know only the dialects still live, so to speak, in relatively separate isolated and traditionally oriented speech communities. Their contacts with the outside world — if we leave aside the intra-local networks — tend to cluster around neighboring towns or urban centers, and even when they do visit these centers they do not usually come in touch with persons of different regional and social origin. Participation of these dialect groups in political life is limited.

Conditions at the regional language level are different. The large Indian cities derive their population from widely divergent sources, including speakers of different sub-regional dialects as well as persons from other regional language areas. The Marwaris of Rajasthan, Sikhs of Punjab and other such communities have large colonies in almost every urban center between Bombay and Calcutta and these colonies maintain regular contacts. Recent anthropological research shows that even among the laboring classes we find groups which migrate thousands of miles in search of work. There is also an increasing group of professionals and government employees who shift from city to city. These are the people among whom we find the greatest percentage of native speakers of the standard. Politically they constitute the active and innovating groups of the population, those among whom new ideas of national independence and religious reform arose and took root. Since Hindi and Urdu act as symbols of these reform movements, villagers and small town residents who aspire to participate in them have gone out of their way to learn the new styles. Although there are no statistics to prove this, it seems reasonable to assume that a high proportion of speakers of the standard are literate. Literacy provides direct access to newspapers and related means of mass

communication. Radio and the cinema are available to urban literates and non literates. All these media help to lend unity to the urban speech stratum and set it apart from the countryside which is as yet little affected.[14] If we were to compare the communication density — measured in terms of personal contact and mass media — between distant urban centers as Patna and Delhi, or Kanpur and Jubbalpur, we might find that it is equal to or even greater than that between geographically less distant areas, say a village in the Awadhi region around Lucknow and another in the Braj area near Aligarh in U. P. There is therefore some justification for speaking of a single urban speech community connected by a super-regional network of communication which extends from one urban center to another, without directly touching the intervening rural areas.

The average individual living within this community does not confine himself to a single style, but has at least some knowledge of all. Literary Hindi and literary Urdu in their extreme forms are cultivated by separate social groups and are considered separate courses of study. Among the educated there are some who read only writings in the Devanagari script; a second group knows only the Perso-Arabic alphabet; a third reads both and others again are comfortable with neither style and do most of their reading in English. Since independence Hindi has become compulsory in schools but Urdu continues to be used extensively in commerce and the Ghazal, the best known form of Urdu poetry, is universally popular. If we look at the modern realist Hindi writers, we find that they utilize both Sanskrit and Persian borrowings. The juxtaposition of the two styles serves to express subtle shades of meaning and to lend reality to their writings. Similarly on the conversational level the use of Hindi and Urdu forms is not simply a matter of birth and education. But, just as it is customary for individuals to alternate between dialect and standard depending on the social occasion, so when using the standard itself the speaker may select from a range of alternatives. Hindi and Urdu therefore might best be characterized not in terms of actual speech, but as norms or ideal behavior in the sociologist's sense.[15] The extent to which a speaker's performance in a particular communication situation approximates the norm is a function of a combination of factors such as

family background, regional origin, education and social attitude and the like.

In our linguistic description we will try to observe the distinction between ideal and actual behavior by comparing our texts collected by ordinary linguistic field methods with normative styles, those features which are regarded as suitable for formal communication and serve as models for actual speech. The fact that we draw this distinction does not mean that we intend to revert to an uncritical acceptance of the traditional prescriptive grammar, but rather we include only those features which seem to us to have a real prestige function. Modern linguists have shied away from the study of prescriptive grammar in recent years on the basis of their experience with Western languages where these grammars too often represent grammatical fictions adhered to by certain groups of grammarians and having no real currency in the community at large. In India this is not the case. Literary styles are learned through many years of oral mimicry memorization — both at home and in the class room — and many of the phonemic distinctions not part of the common core are acquired in this way.

A detailed study along the above lines would of course require extensive field work. Normative styles would have to be validated for each region with questionnaire techniques of the type used by social psychologists and would then have to be compared with an adequate sample of language texts collected under carefully controlled interview conditions. The present paper is a preliminary statement, based on a limited amount of data, which might help to define some of the research problems involved and isolate variables that could be tested in a more detailed study. Information on normative styles was collected by consulting traditional grammars and in consultation with those among our informants who could qualify as experts. The main body of the texts was collected from seven native informants, from different parts of the Hindi regional language area. Informant A is a native speaker of Punjabi who has lived most of his life in Delhi. Informant B is from the Meerut district of U. P., C is from Lucknow, F and G from the Bhojpuri regional dialect area of U. P. D is an additional speaker of Bhojpuri from Bihar and E is a native speaker of Maithili from Bihar. All of them are well educated and have had some formal training

in one or both of the two literary styles. C and F are Muslims trained in Urdu. C has an advanced degree in the language and has also studied some Hindi. G and D are teachers of Hindi and Sanskrit. A, B, and E are graduate students in other subjects with some training in the literary styles.

Except for G and E each informant was subjected to three sets of interview conditions, a) informal conversation with friends, b) reading a passage of prose in a formal style containing features associated with the literary Hindi under conditions approximating those of a lecture to a learned audience, c) reading of a similar passage in literary Urdu. Whenever possible the literary passage was given in the original script, but for informants who did not know the script, passages were put into romanized transcription.

2. The formal and informal styles of Hindi-Urdu differ in phonology, morphology, syntax and lexicon. In this paper we confine our discussion to the most important items in the first two of these categories.

2.1 Maximum Phonemic Inventory:

The table below is a listing of the total range of contrasting segmental phonemes occurring in the data. There are some speakers whose speech shows all these contrasts. The majority however does not have all. Segments not found in all idiolects are placed in parenthesis.

Consonants:

	bilabial	apico-dental	apico post-dental	lamino-palatal	dorso-velar	back dorso-velar
Stops voiceless	p, ph	t, th	ṭ, ṭh	c, ch	k, kh	(ḳ)
voiced	b, bh	d, dh	ḍ, ḍh	j, jh	g, gh	
Spirants voiceless	(f)	s	(ṣ)	(š)	(x)	h
voiced	v	(z)		y	(ġ)	
Nasals	m	n	(ṇ)	~ (nasalization)		
Laterals		l r (trill)	(ṛ) (flap)			

Vowels:

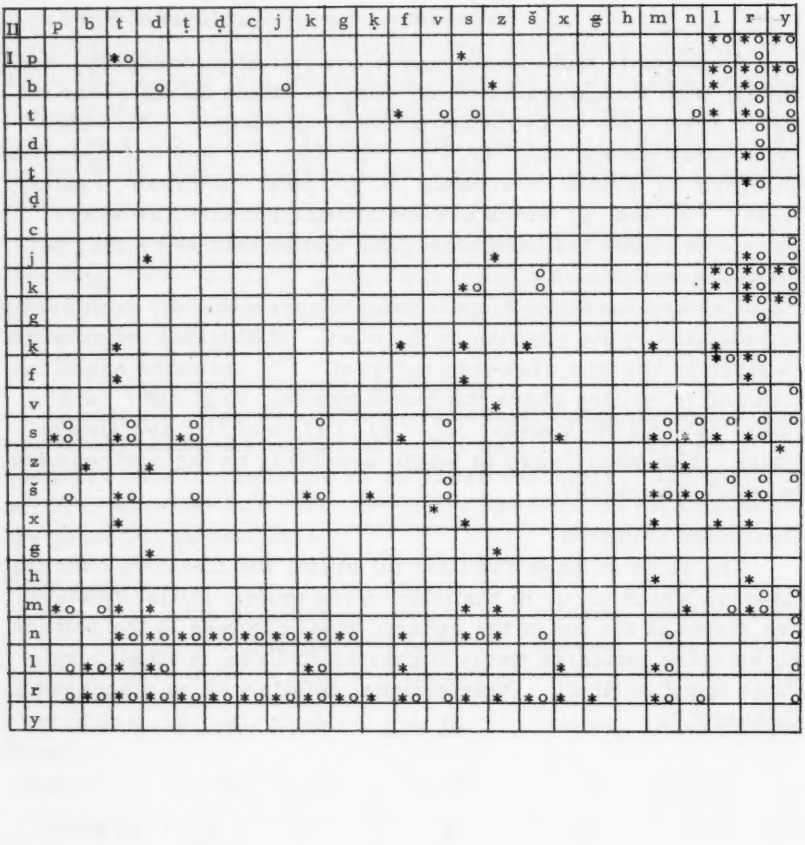
	Front	Central	Back
High	i		u
	I		U
Mid	e	ə	o
	(E)		
Low	æ		ɔ
		a	(ɑ) (unrounded)

Consonants and vowels may occur initially, medially, and finally with the exception of /ṇ/ and /ṛ/ which do not occur initially. Aspirate stops do not occur as geminates. All other consonants except /h/, /x/, /ṛ/, /g̱/, /š/, /ṣ/ and /ṇ/ appear as geminate or double consonants, e. g. /pəta/ 'address', /pətta/ 'leaf'. /v/ and /y/ which are traditionally treated as semivowels are here included among the consonants since they pattern like consonants.

The approximate phonetic range represented by each symbol is indicated by its position in the chart. Additional information is given in the text whenever relevant. /~/ indicates nasalization and is written above the vowel symbol, e. g. /hæ̃/ 'are', /hæ/ 'is'. Of the vowels, /I/, /E/, /ə/, and /U/ are always short. The others vary in length depending on the environment in which they occur.

Consonant Clusters

The table on page 104 lists all initial and final two consonant clusters which occur in the Hindi-Urdu texts. Initial clusters are shown at the top of the square, final clusters at the bottom. An asterisk indicates items occurring in Urdu, a small circle those found in Hindi. Distributional details will be discussed in section 2.2.



2.2 Phonological Sub-Systems.

There are a number of speakers whose speech shows all the contrasts in the maximum inventory; the majority however does not have all of them. If we subtract the segments with limited distribution from the maximum inventory, we arrive at what we may call the common core, representing the minimum number of contrasts necessary for a native speaker of Hindi.

Historically, segments not in the common core did not form part of the local dialects from which the modern standards developed. They came into the system through the influence of other literary languages. /š/ occurs in loan words from Sanskrit (Sk.), Persian (Pe.), Arabic (Ar.), English (En.), and other Western European languages, e. g. /šobha/Sk. 'beauty', /šadi/Pe. 'marriage'. /f/ and /z/ were introduced through Persian, Arabic, English etc., e. g. /fɔrən/Ar. 'immediately', /farm/En. 'farm', /bazar/Pe. 'market', /rezər/En. 'razor'. /ḳ/, /x/, and /ġ/ occur in words of Perso-Arabic sources only, /ḳərib/Ar. 'near', /xUda/Pe. 'god', /ġæb/Ar. 'secret'. /ṇ/ and /ṣ/ are found exclusively in direct borrowings from Sanskrit, e. g. /karəṇ/Sk. 'reason', /kəṣṭ/Sk. 'trouble'. Two additional items occur only in English loan words. These are /E/ and /ɑ/, e. g. /cEk/En. 'check', /bɑl/En. 'ball'. The position of /ṛ/ is somewhat different. The original local dialects seem to have had two phonetic segments [ṛ] and [ḍ] which were in complementary distribution, [ṛ] occurring medially between oral vowels and finally, [ḍ] occurring elsewhere. English borrowings such as /soḍa/ 'soda water' however introduced contrasts between these two allophones and as a result the distinction became phonemic.

Although we refer to the above segments as loans it should be remembered that most of them have been in common use since quite early in the history of the modern standards and by now are regularly part of everyday speech. Their occurrence and non-occurrence however forms a convenient basis for distinguishing sub-systems within the maximum inventory.

The normative Urdu pronunciation may be defined as the pronunciation considered suitable for formal speech making or reading of literary Urdu prose. It consists of the common core plus the following segments: /f š x z ġ ḳ ṛ/. Although normative Urdu is traditionally taught with reference to the writing system it does not incorporate all the distinctions made in the

latter. The letters svad, se, and sin, originally Arabic /ṣ Θ s/ respectively, are all pronounced alike as /s/. zvad, ze, zal and zo which in Arabic represented /ḍ z ḏ ẓ/ are all pronounced /z/. te and to, Arabic /t ṭ/ are pronounced /t/. ain and hamzah which were Arabic /ʕ ʔ/ respectively are no longer pronounced as consonants.

A glottal stop [ʔ] is occasionally heard in the speech of older traditionally trained individuals, e. g. [bəʔid] for /bəid/ 'far' (spelled with ain), but this pronunciation seems to carry little prestige. The pronunciation of Sanskrit loan words is regularly assimilated to the Urdu system Sk. /ṇ/ becomes /n/ and Sk. /ṣ/ becomes /š/, e. g. /karən/ 'because' /bhaša/ 'language'.

The normative Hindi system, the pronunciation considered suitable for formal Hindi, consists of the common core plus: /ṇ/, /š/, /ṛ/ and /ṣ/. Traditional pandits also teach a high central vowel pronunciation [ɨ] in the sequence [rɨ], in words written with the corresponding syllabic r symbol in Sanskrit. But this feature seems to have little real prestige function and might be considered part of a special pandit style. On the level of phonemic distribution, normative Hindi differs from Urdu by the use of final short vowels /I/ and /U/ as in /pətI/ 'husband' and /kIntU/ 'but'. Both styles share final short /ə/ as in /nə/ 'not'. Whereas in normative Urdu consonants in Sanskrit borrowings are generally assimilated, the position of Persian loan segments within the normative Hindi system will require some clarification. Traditional grammars treat these sounds as being outside the system proper and there are some modern writers of Hindi who object to the use of special Devanagari symbols to represent them. In practice /k̲/ /g̲/ and /x/ are generally assimilated, /f/ and /z/ however are quite common at least in Western U. P. and those who do not pronounce them properly are often ridiculed. More detailed sociolinguistic investigation into the prestige position of these items should be of great interest.

The vowel segments /E/ and /ɑ/ as in /cEk/ 'check', /bɑl/ 'ball' are found primarily in the speech of English speaking bilinguals. Their position with respect to the two normative systems is uncertain. Quite possibly more detailed study might isolate a third, Westernized Hindi-Urdu normative style characterized by /E, ɑ/ as well as /z, f, š, r̲/.

In addition to the differences in phonemic inventory and

distribution, normative Hindi and Urdu show a number of important distinctions in consonant clusters. For Urdu our cluster chart lists only a limited number of initial sequences of the type consonant plus l, r or y. Hindi shows many more items in this category and in addition has several other cluster types. Among the more distinctive Hindi initial clusters we might list the following:

kš	kšəma	'forgiveness'
st	stUti	'praise'
sv	svərg	'heaven'
šr	šri	'Mr.' (Urdu /šIri/)
sn	sneh	'affection'
ny	nyay	'justice'

The number of final clusters is larger than that of the initial ones in both styles. Many of these clusters are shared, but each has a few distinctive sequences. Aside from combinations involving loan phonemes, the following clusters might be considered as distinctively Urdu. a) Combinations of m plus non-homorganic stop or spirant: mt, md, ms, mn. b) Sequences with final l; Hindi has only ml in this category, Urdu has bl, tl, kl, sl. c) Clusters with final s: Hindi has only ts, Urdu has ps, rs, ms. d) h plus consonant, hr, hm:

mt	sImt	'direction'
ms	ləms	'touch'
kl	šəkl	'shape'
hr	šæhr	'city' (Hindi /šəher/ or /šəhər/

Most important among the distinctive final Hindi clusters are sequences of consonant plus v and y. These do not occur in Urdu:

tv	məhətv	'importance'
šv	əšv	'horse'
ny	əny	'other'
ly	muly	'price'
rv	gərv	'pride'
jy	rajy	'rule' (Urdu raj)
ry	sury	'sun'

It is interesting to note that many of the sequences with final /y/ are re-sanskritizations of borrowings which in earlier Hindi texts appear with simple consonant endings. Thus Bharatendu Harischandra, the most important of the early Hindi writers, uses /raj/ not /rajy/ 'rule'.[16]

The two styles further differ with respect to clusters involving aspirate stops, which were not listed in the chart. In Urdu aspirate stops occur only in final clusters after nasals. Hindi has the following types of additional sequences:

rth	ərth	'meaning'
rbh	gərbh	'embryo'
rkh	murkh	'fool'
khy	mUkhy	'chief'

2.3 Observed pronunciations.

In this section the normative pronunciations are compared with field data from India and the formal and informal interview texts collected for this study. Differences among informants are classified according to the level of structure at which they appear, i. e. according to whether they affect the phonemic inventory, phonemic distribution or whether they are phonetic.

Differences in Phonemic Inventory.

The majority of the uneducated illiterate speakers of Hindi and the regional dialects ordinarily replace the borrowed segments /š/, /f/, /x/, /z/, /ġ/, and /ṇ/ by /s/, /ph/, /kh/, /j/, /g/, and /n/ respectively, and /ḍ/ and /ṛ/ for them are allophones of the same phoneme. With the educated informantss recorded for this study the situation is more complex. All of them regularly distinguish between /š/ and /s/. The segment /ḳ/ occurs only with informants C and F, both trained in Urdu. All others do not have it. The most common reflexes found in words with an original /ḳ/ were /x/ before voiceless stop and /k/ elsewhere, e. g. /kanun/ for /ḳanun/ 'law', /vəxt/ for /vəḳt/ 'time'.

Informants C and F also regularly observe the contrast between /ġ/ and /g/, e. g. /gUl/ 'noise', /gUl/ 'flower'. E uses /g/ in all words of this type, and thus has no phoneme /ġ/. With A and B the position is as follows. There are some items in which the pronunciation varies between [ġ] and [g], e. g. [ġələt]~[gələt] 'wrong'. Others such as [gəŋga]

'Ganges' always have [g]. The phonetic alternation is confined to a limited list of items and is not predictable in terms of the phonological environment. It is necessary therefore to set up /g/ as part of the phonemic inventory of informants A and B, although it is marginal. Items like the word for 'wrong' above are said to have two phonemic forms: we write /gələt/ when [g] is heard and /gələt/ when we hear [g].

The segment /x/ regularly occurs with informants A, B, C, and F. D and E show variations similar to the /g/ and /g/ alternation described above, e. g. [xas]~[khas] 'special'. /x/ is therefore phonemic in their speech.

Similar alternations between [j] and [z], and [ph] and [f] are found with informants B, D, and E, e. g. [cij]~[ciz] 'thing', [kaphi]~[kafi] 'enough'. The distinctions between [j] and [z] and [ph] and [f] are therefore phonemic for them.

The contrast between /n/ and /ṇ/ is regularly found only with E, B, A, and D. Informant C varies between [n] and [ṇ] in certain words. /ṇ/ is a marginal phoneme in his dialect. F never shows [ṇ]. The contrast between /š/ and /ṣ/ was heard only with informant E and his speech shows variation between [š] and [ṣ] in words with Sk. /ṣ/.

An interesting feature of the alternations between 'loan' and 'native' phonemes is their apparent irregularity in certain loan words and their regularity in others. Informant E for example regularly has /j/ in /jIla/ 'district' and /jərur/ 'certainly', but alternates between [z] and [j] in other items like /zyada/ 'more'. The consistent use of either of these two phonemes in any particular items suggests that these items were borrowed at a quite early stage directly into the regional dialects. The /f/ and /ph/ alternation for some informants extends also to certain words which are Indic in origin and should have /ph/ consistently, e. g. /phIr/~/fIr/ 'again', /phəl/~/fəl/ 'fruit'.[17]

Differences in Phonemic Distribution.

The most important difference in this category is in the distribution of vowels. Informants D, E, have contrasts between /a/, /i/, /u/ and /ə/, /I/, /U/ in final position, e. g. /nə/ 'no' /ka/ 'of' /yədI/ 'if' /nədi/ 'river', /kIntU/ 'but' /babu/ 'clerk'. Other informants have final /ə/ and /a/ but only one high front and high back vowel /i/ and /u/ respectively in this position. D and E further have the final sequence

/əh/ in a number of polysyllabic words including words of both Indo-Aryan and Perso-Arabic origin, e. g. /barəh/ 'twelve', /terəh/ 'thirteen', /jəgəh/ Per. 'place', /tərəh/ Ar. 'kind'. The remaining informants seem not to have this final /əh/. Correspondences vary between /a/ or /ha/ in words of IA origin like the numerals and further also include /e/ or /he/ in some of the loan words, /bara/~/barha/ 'twelve', /jəga/~/jəgha/~/jəge/~/jəghe/ 'place'.

There are other alternations in phonemic distribution of varying degree of importance which cannot be discussed here in any detail but which might be worthy of further investigations. One might mention the variation between single and double consonants in items like /buḍha/~/bUḍḍha/ 'old'. Similarly of interest are the alternations between /kš/ and /kkh/, e. g. /dəkšIṇ/~/dəkkhIṇ/ 'south', between /v/ and /b/, e. g. /vən/~/bən/ 'forest', between nasalized and non-nasalized vowels, e. g. /ĩṭ/~/iṭ/ 'brick', and many more.

Phonetic Differences.

On the phonetic level the following variations were found. With informant A the phoneme /ə/ has allophones ranging between [ɨ] and [ə] before certain retroflex consonants. Other informants vary between [ə] and [ʌ] in those environments. Allophones of /ə/ before /h/ cluster around [æ] or [ɛ] for most informants. E and D have [ʌ] in this environment. The allophones of /i/, /I/, /e/, /u/, /U/, and /o/ are comparatively high and tense with D, E, and F. They are lower and more lax with A, B, and C.

/ɔ/ and /æ/ represent monophthongal low back and front vowels in the case of informant A. With B and C the pronunciation is diphthongal with a low vowel onset followed by a very short upward glide [ɛe] and [ɔo]. The diphthongal glide with a central vowel onset, [əɛ] and [əo], is pronounced with informants D, E, and F.

/y/ and /v/ are pronounced with relatively strong consonantal friction by A, B, and C. Informants D, E, and F have a great deal more of vowel color.

The tongue position of the retroflex consonants /ṭ/ and /ḍ/ is most retracted and the preceding vowel has most retroflex color with informant A. Informants B, and C have less retracted tongue position and D, E, and F have least retroflection.

The progressive loss of phonetic distinctiveness of the dental-retroflex contrasts is of interest in view of the fact that the dialects of Bihar are closest to Assamese, which does not have this contrast.

Analysis of the pronunciation differences observed so far shows that the occurrence and non-occurrence of borrowed segments is related both to education or family background and to regional origin. The two Moslems C and F who have had formal training in Urdu agree in the consistent use of Perso-Arabic loan segments. All Hindus and C who has had some Hindi training have /ṇ/. With informants from the West /z/, /f/, and /x/ are regular parts of the phonemic inventory, whereas with D and E from Bihar these contrasts are marginal. This seems to reflect the fact that Indo-Islamic influence has been most profound in the West and Urdu literature most widely accepted. Differences on the phonetic level on the other hand are correlated primarily with regional origin. Informant A who has lived in Delhi most of his life is set off from all others by his pronunciation of /ə/, /æ/, /ɔ/, /ṭ/ and by forms such as /terha/ 'thirteen', features which seem characteristic of the sub-regional dialects in the Hindi-Punjabi border area. Similarly D, E and F, the two Hindus and the Moslem from Bihar show similarities in the pronunciation of vowels, of /ṭ/ and /ḍ/, /y/ and /v/ and of final /h/. All these features are also found in the sub-regional dialects of Bihar.

The data regarding situational variants in pronunciation is meager. Informant C uses /ṇ/ only in situations calling for normative Hindi. The speech of informant D shows a much higher incidence of /z/ and /f/ in conversations with Moslems and Westerners; /j/ and /ph/ are more prevalent when he talks to his fellow Hindu from Bihar. Both authors have observed additional situational variants in India with older Hindus, trained in both styles. These individuals use /k̤/, and /g̤/ only in situations which call for normative Urdu. With other informants interviewed for this study variations of this type were less pronounced. More detailed investigation is needed with closer specification and better control of interview conditions.

There is furthermore little correlation between formal and informal interview situations and the pronunciation differences discussed so far. Formal-informal differences are more

clearly evident in the pronunciation of consonant clusters. Uneducated low prestige speech shows the smallest number of clusters. In this style only sequences of consonant plus /y/ occur initially and finally there are only clusters of nasal plus homorganic stop or sibilant plus stop. Clusters in common loan words are replaced as illustrated in the patterns below:

/glas/ 'glass'	/gIlas/
/skul/ 'school'	/Iskul/ (East) or /səkul/ (West)
/kšəma/ 'forgiveness'	/chəma/ or /šəma/
/xətm/ 'end'	/xətəm/
/šrəm/ 'labor'	/šərəm/
/šərm/ 'shame'	/šərəm/
/nəgd/ 'cash'	/nəkəd/ or /nəgəd/ or /nəgd/
/rəsm/ 'tradition'	/rəsəm/
/šUkr/ 'thanks'	/šUkUr/ or /šUkər/
/prem/ 'love'	/pIrem/ or /pərem/
/rajy/ 'rule'	/raj/
/smərən̤/ 'remembrance'	/sUmirən/

Educated speakers regularly use initial clusters of the types Cr, Cy, Cl, and Cv, except for occasional variations in individual words such as that between /pleṭ/ and /pəleṭ/ 'plate'. Initial sC sequences however tend to be replaced by VsC in informal contexts, e. g. /skul/~/Iskul/ 'school', /sṭešən/~/Isṭešən/ 'station'. In final position similar formal-informal alternations are found in /k̤Ism/~/k̤IsIm/ 'kind'. /bəks/~/bəkəs/ 'box' /jənm/~/jənəm/ 'birth'. With a number of final clusters, alternations with consonant vowel consonant sequences are possible in some words but not in others. Thus educated speakers may say /bərəf/ or /bərf/ 'snow', but always use /sIrf/ 'only'. Similar examples are: /təxt/~/təxət/ 'throne, bench' versus /səxt/ 'hard'; /kIlk/~/kIlIk/ 'reed' versus /mUlk/ 'country'; /bhəsm/ /bhəsəm/ 'ashes' versus /jIsm/ 'body'; /muft/~/mUfət/ 'free' versus /koft/ 'disgust'; /məkr/~/məkkər/ 'deceit' versus /šUkr/ 'thanks'. A third set of items always shows final clusters: /šəbd/ 'word', /šəxs/ 'person', /svərg/ 'heaven'.

Clusters not shared in both normative styles tend to be replaced in the style in which they do not occur. Some common instances are: U. /šIri/ for H. /šri/, U. /raj/ for H.

/rajy/ 'rule', U. /krIšən/ for H. /krIšṇ/ 'Krishna', H. /šəher/ for U. /šæhr/.

2.4 Morphological Differences.

A complete analysis of morphological variants within Hindi-Urdu is beyond the scope of this paper. We will note here only the most striking differences.

One major distinction is found with demonstrative pronouns. Normative Hindi has four different forms, which are kept distinct in spelling and pronunciation:

/yəh/ 'this', /ye/ 'these' /vəh/ 'that', /ve/ 'those'

Normative Urdu has only two forms, written yəh, vəh but pronounced as below:

/ye/ 'this, these' /vo/ 'that, those'

Informants A and F always use /ye/ and /vo/, C employs the first set only in formal Hindi contexts. B, D and E alternate between /yəh/ and /ye/ and B also changes between /vəh/ and /vo/ in the singular. All three keep the plural /ve/ distinct. The Hindi normative set is most closely adhered to in formal situations. It would also seem that the Urdu system is more common in the West, but more information is needed on the details of the alternation and on its social function.

Borrowings from Persian and Sanskrit, in addition to creating differences in lexicon, have also brought in a number of differences in the inventory of derivational morphemes. Thus in Urdu we have:

-ən	/nIsbətən/ 'relatively'
-in	/rəngin/ 'colorful'
-mənd	/gərəzmənd/ 'desirous'
-Iya	/məzaḳIya/ 'funny'
həm-	/həmdərd/ 'friend' (lit. one who shares the pain)
ba-	/baḳaIda/ 'according to rules'
-nak	/xətərnak/ 'dangerous'
pUr-	/pUrlUtf/ 'pleasant'
dər-	/dərmIan/ 'middle'
-šInas	/fərzšInas/ 'conscientious, dutiful'
bəd-	/bədnam/ 'notorious'

Some characteristic Hindi affixes are:

-təh	/səmbhəvətəh/ 'probably'
-lu	/šrəddhalu/ 'respectful'
-ək	/səhayək/ 'helpful'
-akar	/golakar/ 'circular'
-Ik	/pərIvarIk/ 'domestic, of family'
dUr-	/dUrdəša/ 'bad condition'
-ta	/kəṭhorta/ 'brutality'
ən-	/ənərth/ 'useless'
svə-	/svəbhav/ 'nature'
prə-	/prəkop/ 'extreme anger'
ə-	/əvak/ 'silent'
-man	/bUddhIman/ 'wise'
-It	/jivIt/ 'alive'
Ud-	/Uddhar/ 'release'
-purvək	/bhəktIpurvək/ 'full of piety, faithfully'
-tv	/məhətv/ 'responsibility'

It is interesting to note that the bulk of the Sanskrit loan affixes are freely used only with direct Sanskrit borrowings (tatsamas) and the Perso-Arabic loan affixes are used only with Perso-Arabic borrowings. Only a few have come to be used also with indigenous roots (tadbhavas). Among these are:

-dar	/dukandar/ 'shopkeeper', /ṭhekedar/ 'contractor'
be-	/bejan/ 'lifeless', /bedhəṛək/ 'fearless'
-ana	/bUzUrgana/ 'elderly', /jogiana/ 'like a jogi'
na-	/nagəvar/ 'unbearable', /nasəməjh/ 'unwise'
-xana	/dəvaxana/ 'hospital', /kuṛaxana/ 'garbage pile'
ən-	/ənpəḍh/ 'illiterate', /ənərth/ 'meaningless'

The first five items in the above list are of Persian origin and are used with Persian loans and tadbhavas. The last item is from Sanskrit and is used with tatsamas and tadbhavas. We have no instance of a Persian loan affix used with tatsamas or a Sanskrit loan affix used with Persian loan roots.

In normative Hindi a number of roots undergo morphophonemic changes on the analogy of Sanskrit Sandhi, e. g. /ved/ 'Veda', /vædIk/ 'Vedic'. Traditional prescriptive grammars as a rule describe these changes in terms of the Sanskrit system, using the same classification and terminology.[18] In

practice the Hindi system is somewhat different from the Sanskrit —we may have for example both /rajnitIk/ and /rajnætIk/ 'political'— but these differences have not as yet been described.

As with consonant clusters, the speech of the uneducated shows the smallest number of derivative compounds. When arranged in order of formality the affixes roughly fall into three groups. The items in the third set of examples given above and few others such as -Iya, -bəd are in common use. Others indicate either relatively formal Urdu, e. g. -mənd, ba- or relatively formal Hindi, e. g. -lu, -ək, -akar, -ta. A third group of very formal items is productive primarily in writing and very rarely in speech, e. g. Urdu -ən, pUr- or Hindi -təh, -Ik, -purvək-, -tv. In ordinary speech the relatively formal and very formal affixes are most usually substituted by phrases. Thus /səmbhəvətəh/ 'probably' becomes /səmbhəv hæ/ or /mUmkIn hæ/ or /nIsbətən/ 'relatively' becomes /Us ke mUḳable mẽ/ etc.

Among the lexical borrowings from Persian and Sanskrit there are also a number of common conjunctions and adverbs which have considerable structural importance, and might be classed as function words.[19] These form another important criterion for stylistic distinctions. The listing below gives items occurring in normative Hindi, normative Urdu and informal conversation.

Normative Hindi	Informal Conversation	Normative Urdu
/yədI/, /əgər/ 'if'	/əgər/	/əgər/
/yədyəpI/ 'although'	/əgər/	/əgərcI/, /halākI/
/təthapI/, /to bhi/ /phIr bhi/ 'yet', 'nevertheless'	/phIr bhi/, /to bhi/ /təb bhi/	/təb bhi/, /phIr bhi/
/evəm/, /tətha/, /ɔr/ 'and'	/ɔr/	/ɔr/
/karəṇ/ 'reason'	/karən/, /vəja se/ /vəje se/	/səbəb se/, /vəje se/
/kIntU/, /pərəntU/ /məgər/ 'but'	/məgər/, /lekIn/	/məgər/, /lekIn/
/sərvda/, /səda/, /nIty/ 'always'	/səda/, /həmeša/	/səda/, /həmeša/
/səhIt/, /sath/ 'with'	/sath/	/sath/
/ətəh/, /Is lIe/ 'therefore'	/Is lIe/	/cUnancI/, /Is lIe/

The items in the normative Hindi and Urdu columns can again be arranged into relatively formal and very formal groups. Items in the first category are: Urdu /halãkI/ 'although', /cUnancI/ 'therefore' and Hindi /yədI/ 'if' /pərəntU/ 'but'. In the latter group we find Hindi /təthapI/ 'yet', /evəm/, /tətha/ 'and' /kIntU/ 'but' /səhIt/ 'with', /ətəh/ 'therefore'. It is interesting to note that none of the function words in the normative Urdu column are in the very formal group. A characteristic feature of very formal Urdu style is the use of the enclitic -e 'of' and -o 'and' borrowed from Persian, e. g. /sahəbe xana/ for the more common /ghar ka sahəb/ 'owner of the house', /dərde dIl/ for /dIl ka dərd/ 'pain of the heart', /abo həva/ for /həva aur pani/ 'climate' (lit. 'air and water'), /səvalo jəvab/ for /səval aur jəvab/ 'question and answer'.

3. Conclusion.

The standard speech stratum in the Hindi regional language area can be described in terms of a single range of styles. Within this range we may distinguish a common core, shared by all including the uneducated and two, or possibly three, formal extremes or normative styles which are set off from each other by features of phonology, morphology and syntax. In actual texts, spoken as well as written, the differences between the extremes tend to be blurred somewhat but we may distinguish various degrees of Hindi, Urdu and Westernized coloring. The various styles within the standard stratum tend to be associated with particular communication roles and the speech of any one individual varies accordingly. Some individuals are at home in both normative styles, with others the variations are more limited. Aside from the Hindi-Urdu differences, we find speech differences such as the contrast between /s/ and /š/ and the use of certain consonant clusters, which correlate with differences in education. Degree of formality and informality is signalled by other consonant clusters, by derivative suffixes, borrowed enclitics and certain types of function words.

The exact nature and social function of the above stylistic markers varies from region to region. In the West for example the contrast between /z/ and /j/ is a regular part of the phonemic inventory of educated speakers; in the East it is marginal for many and only persons with special training in

Urdu have it regularly. Regional origin is most clearly evidenced on the phonetic level. All speakers from a particular area tend to have the same allophones regardless of social background. A person from Bihar trained in Urdu might have the same phonemic inventory as another Urdu speaker from Lucknow, however, in items which are part of the common core, the allophonic content of his speech resembles that of other speakers from the same area who often have different phonemic inventories.

Notes

1. For a somewhat different and more detailed discussion of the general Hindi or Hindustani problem see S. K. Chatterji, Indo-Aryan and Hindi (Ahmedabad, 1942).

2. The variations within Hindi are similar to the differences between Colit Bhasha and Sadhu Bhasha in Bengali, described by Dimock and Chowdhury in this volume.

3. See for example the following: S. H. Kellog, A Grammar of the Hindi Language (3rd ed., London, 1938); K. P. Guru, Hindi Vyākaran (Banaras, 1957); A. Sharma, A Basic Grammar of Modern Hindi (Delhi, 1958), this is the most recent official grammar published by the Government of India, Ministry of Education and Scientific Research.

4. The standard reference work is John T. Platts, A Grammar of the Hindustani or Urdu Language (London, 1920).

5. See for example A. H. Harley, Colloquial Hindustani (London, 1944); M. C. Saihgal, Hindustani Grammar (11th ed., Simla, 1945); T. G. Bailey, Teach Yourself Hindustani (London, 1950). Some of the grammars in our three categories discuss both scripts but do not discuss the relations between the styles.

6. Sir George A. Grierson, Linguistic Survey of India (19 vols., Calcutta, 1903-28).

7. The most useful of the traditional treatments is S. K. Chatterji, op. cit., but this contains little concrete linguistic data.

8. Cf. J. J. Gumperz, "Language Problems in the Rural Development of North India," Journal of Asian Studies, 16. 251-9 (1957). For a discussion of similar local dialect chains elsewhere; cf. Trevor Hill, "Institutional Linguistics," Orbis, Tome VII, No. 2, 1958, pp. 441-455. See also the discussion

of L complex and L simplex in C. F. Hockett, A Course in Modern Linguistics (New York, 1955) p. 323 ff.

9. Sir George A. Grierson, op. cit.

10. Cf. O. H. K. Spate, India and Pakistan (London, 1954), the standard geography of India which lists Bihari and Western Hindi along with Gujerati, Bengali, etc.

11. S. K. Chatterji, who suggests that Bazaar Hindi be made the national language, does not consider these prestige relationships.

12. Cf. R. Redfield, Peasant Society and Culture (Chicago, 1956), p. 70 ff.

13. For a general discussion of relevant communication problems see K. W. Deutsch, Nationalism and Social Communication (New York, 1953).

14. Cf. Deutsch's concept of mobilized population, ibid., p. 99 ff.

15. Cf. K. Davis, Human Society (New York, 1948), p. 55 ff.

16. Cf. R. C. Shukla, Hindī Sāhitya Kā Itihās (Banaras, 1955), p. 529.

17. Cf. the change from ph to f in 19th century Bengali described in S. K. Chatterji, Origin and Development of the Bengali Language (Calcutta, 1926), Vol. I, p. 442.

18. Cf. K. P. Guru, op. cit.

19. For the use of this term see C. C. Fries, The Structure of English (New York, 1952).

INDIANA UNIVERSITY PUBLICATIONS IN ANTHROPOLOGY AND LINGUISTICS

1. Paul Radin, Winnebago Hero Cycles: A Study in Aboriginal Literature (168 pp., 1948) $2.00
2. Paul Radin, The Culture of the Winnebago: As Described by Themselves (119 pp., 1949) $2.00
3. Paul Radin, The Origin Myth of the Medicine Rite: Three Versions. The Historical Origins of the Medicine Rite (78 pp., 1950) $2.00

4–5. Harold E. Driver and S. H. Riesenberg, Hoof Rattles and Girls' Puberty Rites in North and South America (33 pp., 1950); and Hilda J. Curry, Negative Painted Pottery of Angel Mounds Site and its Distribution in the New World (57 pp., 1950) $1.00

6. L. S. Freeland, Language of Sierra Miwok (205 pp., 1951) $3.00
8. Claude Lévi-Strauss, Roman Jakobson, C. F. Voegelin and Thomas A. Sebeok, Results of the Conference of Anthropologists and Linguists (86 pp., 1953) $1.00
9. Harold E. Driver, John M. Cooper, Paul Kirchhoff, Dorothy Rainier Libby, William C. Massey, and Leslie Spier, Indian Tribes of North America (Map + 30 pp., 1953) $1.00. Outline maps 11 x 17 inches may be purchased separately
11. Charles F. Hockett, A Manual of Phonology (246 pp., 1955) $3.50
13. Eric H. Lenneberg and John M. Roberts, The Language of Experience (33 pp., 1956) $1.00
14. Charles F. Voegelin and Florence M. Voegelin, Hopi Domains: A Lexical Approach to the Problem of Selection (82 pp., 1957) $2.00
15. Sarah C. Gudschinsky, Proto-Popotecan: A Comparative Study of Popolocan and Mixtecan (118 pp., 1959) $2.00
16. George William Grace, The Position of the Polynesian Languages Within the Austronesian Language Family (77 pp., + Map, 1959) $2.00

Orders for Memoirs may be sent to International Journal of American Linguistics, Department of Anthropology, Indiana University, Bloomington, Indiana

PUBLICATIONS OF THE RESEARCH CENTER IN ANTHROPOLOGY, FOLKLORE, AND LINGUISTICS

1. Edward Sapir and Morris Swadesh, Native Accounts of Nootka Ethnography (457 pp., 1955) $5.00
2. Eeva K. Minn, Studies in Cheremis, Volume 4, Derivation (viii + 99 pp., 1956) $2.00
3. Robert N. Pehrson, The Bilateral Network of Social Relations in Könkämä Lapp District (xiv + 128 pp., 1957) $2.50
4. Leonard Newmark, Structural Grammar of Albanian (vi + 130 pp., 1957) $2.50
5. Robert E. Longacre, Proto-Mixtecan (viii + 195 pp., 1957) $3.50
6. Stanley Newman, Zuni Dictionary (vi + 117 pp., 1958) $2.50
7. Robert William Albright, The International Phonetic Alphabet: Its Backgrounds and Development (viii + 78 pp., 1958) $2.50
8. Melville Jacobs, Clackamas Chinook Texts, Part I (vi + 293 pp., 1958) $4.00
9. Henry and Renée Kahane and Sol Saporta, Development of Verbal Categories in Child Language (vi + 65 pp., 1958) $2.50
10. V. Propp, Morphology of the Folktale (x + 134 pp., 1958) $5.00
11. Melville Jacobs, Clackamas Chinook Texts, Part II (viii + 369 pp., 1959) $4.00
12. Robert B. Lees, The Grammar of English Nominalizations (xxvi + 206 pp., 1960) $.400
13. Charles A. Ferguson and John J. Gumperz, Eds., Linguistic Diversity in South Asia: Studies in Regional, Social and Functional Variation (viii + 118 pp., 1960) $3.50

Orders will be received by Thomas A. Sebeok, Director of Publications of the Research Center in Anthropology, Folklore, and Linguistics, Indiana University, Bloomington, Indiana